

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

INTEROPÉRABILITÉ DES ÉLÉMENTS DE MÉTADONNÉES:  
VERS UNE APPROCHE SÉMANTIQUE

MÉMOIRE  
PRÉSENTÉ  
COMME EXIGENCE PARTIELLE  
DE LA MAÎTRISE EN INFORMATIQUE

PAR  
BERNARD BIZIMANA

JUILLET 2009

UNIVERSITÉ DU QUÉBEC À MONTRÉAL  
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

Ce mémoire est dédié à mon épouse,

Marie d'Amour Uwumuremyi

à mes quatre enfants,

Carmen Kakuze Bizimana

Éloi d'Amour Bizimana

Élyse Bizimana

Anaïs-Placidia Bizimana

et à ma mère,

Bellancilla Kakuze

À la mémoire de mon père, Laurien Rwabuhungu

À la mémoire du Frère Mansuy Willy Jongen

À la mémoire du Frère Jacobs Ian Paul

## Remerciements

Tout travail de recherche est nécessairement une œuvre collective, et à fortiori, un mémoire de maîtrise comme celui-ci. J'ai donc l'agréable devoir d'exprimer ma profonde gratitude à tous ceux et celles qui, de près ou de loin, ont contribué à sa réalisation.

Tout d'abord, je remercie particulièrement mon directeur de recherche, monsieur Gilles Gauthier, qui a bien voulu diriger ce mémoire et me faire profiter de sa rigueur scientifique. Ses commentaires, remarques et suggestions pertinentes ont grandement enrichi mes réflexions. C'est lors de ses cours sur les technologies XML que j'ai commencé à m'intéresser à la problématique des métadonnées et du Web sémantique. Qu'il soit assuré de ma profonde reconnaissance. Ma reconnaissance va aussi à tous mes collègues de travail qui n'ont jamais cessé de m'encourager dans cette longue et fastidieuse aventure. Merci à tous mes amis qui n'ont cessé de m'encourager tout au long de ces dernières années.

Je tiens à exprimer ma profonde gratitude aux Frères des Écoles Chrétiennes de Byumba, au Rwanda, qui représentent pour moi une source d'inspiration profonde et fortifient depuis mon jeune âge ma formation humaine. Je remercie particulièrement les frères Mansuy et Ian, aujourd'hui décédés, ainsi que le Frère Aloys Ndimukihe.

Je tiens également à remercier un ami de longue date, Pontien Mbaraga, qui m'a accordé tout au long de mes études en informatique, une aide précieuse chaque fois que son expertise en mathématiques était requise. Mes remerciements vont aussi à Rita pour sa générosité et dont les commentaires ont contribué à améliorer la qualité de ce manuscrit. Merci aussi à madame Lise Arsenault,

assistante à la gestion des programmes d'études avancées au Département d'informatique de l'UQAM pour la qualité exceptionnelle de ses services.

Enfin, je tiens à remercier de tout cœur ma famille qui m'a donné l'appui et l'encouragement moral indispensables à la réalisation de ce mémoire. À mon épouse et à mes enfants, merci pour toute votre compréhension tout au long de cette aventure.

## TABLE DES MATIÈRES

LISTE DES FIGURES .....	ix
LISTE DES SIGLES ET ABRÉVIATIONS .....	xi
LISTE DES ANNEXES .....	xii
RÉSUMÉ .....	xiii
INTRODUCTION .....	1
PREMIÈRE PARTIE	
INTEROPÉRABILITÉ DES ÉLÉMENTS DE MÉTADONNÉES : ÉLÉMENTS DE SOLUTION .....	8
CHAPITRE I	
PROBLÉMATIQUE .....	9
1.1. Définition des concepts de base .....	9
1.2. Les différents niveaux d'interopérabilité .....	11
1.3. Les facteurs d'interopérabilité des métadonnées .....	12
1.3.1. <i>Multiplicité des SRI</i> .....	14
1.3.2. <i>Multiplicité des protocoles</i> .....	15
1.3.3. <i>Multiplicité des schémas de métadonnées</i> .....	16
1.3.4. <i>Diversité des langues</i> .....	17
1.3.5. <i>Diversité de vocabulaires</i> .....	18
1.3.6. <i>Diversité des jeux de caractères</i> .....	19
1.3.7. <i>Hétérogénéité des formats de représentation des métadonnées</i> .....	19
1.3.8. <i>Diversité des modèles abstraits</i> .....	20
1.3.9. <i>Mécanismes d'identification des éléments de métadonnées</i> .....	21
CHAPITRE II	
SOLUTIONS ACTUELLES AU PROBLÈME DE L'INTEROPÉRABILITÉ DE MÉTADONNÉES .....	26
2.1. Les méthodes d'interopérabilité au niveau du schéma .....	27
2.1.1. <i>La dérivation</i> .....	28
2.1.2. <i>La traduction</i> .....	28
2.1.3. <i>L'adaptation et les profils d'application</i> .....	29
2.1.4. <i>Le mapping</i> .....	31

2.1.5. Le « <i>switching</i> » .....	34
2.1.6. Utilisation des modèles de référence .....	35
2.2. L'interopérabilité des métadonnées au niveau des instances de métadonnées .....	41
2.2.1. La conversion .....	41
2.2.2. Intégration des métadonnées .....	43
2.2.3. L'interopérabilité au niveau des dépôts .....	43
2.3. Interopérabilité au niveau de la valeur des éléments .....	45
2.4. Conclusion .....	49
CHAPITRE III	
LE WEB SÉMANTIQUE ET L'INTEROPÉRABILITÉ DES MÉTADONNÉES .....	52
3.1. Le Web sémantique .....	53
3.2. Pourquoi le Web sémantique .....	54
3.3. Les technologies du Web sémantique et l'interopérabilité .....	57
3.3.1. Couche de base .....	60
3.3.2. Couche XML .....	61
3.3.3. Couche RDF .....	64
3.3.4. Les ontologies .....	69
3.4. Le Web sémantique : fiction ou réalité ? .....	74
3.5. Vers le Web sémantique : les formats de transition .....	76
3.5.1. Les microformats .....	77
3.5.2. Les attributs RDF .....	78
3.6. Le Web sémantique : perspectives .....	84
DEUXIÈME PARTIE	
DISTRIBUTION DES MÉTADONNÉES DANS LE WEB SÉMANTIQUE: LE CAS DU MLR .....	89
Introduction .....	90
CHAPITRE IV	
METADATA FOR LEARNING RESOURCES .....	93
4.1. Pourquoi le MLR .....	93
4.2. Présentation du MLR .....	96
4.2.1. Le Framework .....	98
4.2.2. Les éléments de base du MLR .....	103



## CHAPITRE V

LE MLR ET LE WEB SÉMANTIQUE.....	106
5.1. Atouts du MLR pour le Web sémantique .....	106
5.2. Intégration des éléments de métadonnées <i>MLR</i> dans le Web sémantique.....	110
5.2.1. Contextes d'utilisation du MLR.....	111
5.2.2 Approche de distribution des notices de métadonnées MLR .....	114
5.2.3 Contributions distribuées à l'indexation et le MLR.....	120

## CHAPITRE VI

CONCLUSION .....	126
BIBLIOGRAPHIE.....	133
ANNEXE 1.....	142

## LISTE DES FIGURES

### Introduction

0.1 :	Rapport entre le rappel et la précision	2
-------	---	---

### Chapitre 1

1.1 :	Niveaux d'interopérabilité des métadonnées	11
1.2 :	Facteurs d'interopérabilité des métadonnées	13
1.3 :	Multiplicité des systèmes de repérage de l'information	14
1.4 :	Identification des éléments de métadonnées	21
1.5 :	Problèmes d'interopérabilité des métadonnées	23

### Chapitre 2

2.1 :	Profil d'application	29
2.2 :	Technique de "switching"	34
2.3 :	Métamodèle de haut niveau de la norme ISO 11179-3	37
2.4 :	Schéma de métadonnées : niveau conceptuel	38
2.5 :	Schéma de métadonnées : niveau de métadonnées conceptuel	39
2.6 :	Conversion des notices issues des schémas différents	42
2.7 :	Problème de recherche <i>sujet</i> dans un contexte multilingue	45
2.8 :	Recherche <i>sujet</i> et barrières linguistiques	46
2.9 :	Cohabitation des vedettes matières dans une même notice	48
2.10 :	Utilisation d'un langage intermédiaire	48

### Chapitre 3

3.1 :	Technologies du Web sémantique et interopérabilité	52
3.2 :	Représentation de l'information pour le Web sémantique	53
3.3 :	Productivité de la recherche par mots-clés sur le Web	55
3.4 :	Couches de base du Web sémantique selon Hori	57
3.5 :	Modèle conceptuel du Web sémantique selon Geroimenko	58
3.6 :	Le Web sémantique selon T. Berners-Lee	59
3.7 :	Évolution des langages de balisage	62
3.8 :	Principales technologies de la famille XML	62
3.9 :	Le modèle RDF	65
3.10 :	Exemple d'un schéma RDF	68
3.11 :	Vocabulaires et richesse sémantique	70
3.12 :	L'ontologie selon Gruber	71
3.13 :	Outils d'annotation et coûts associés	75

3.14 :	Métadonnées internes et externes	79
3.15 :	Principes des balises RDF embarquées	80
3.16 :	Fonctionnement de GRDDL	82
3.17 :	Convergence entre microformats et RDFa	83
3.18 :	Hypertexte et Web sémantique	84
3.19 :	Développement du Web sémantique	85
3.20 :	Évolution du Web selon N. Spivack	86
3.21 :	Courbe en S de Richard N. Foster adapté au Web sémantique	87
3.22 :	Outils du Web sémantique en développement	88

## Chapitre 4

4.1 :	Cadre de métadonnée d'une ressource selon le MLR	96
4.2 :	Carte conceptuelle d'une spécification d'un élément de métadonnée	99
4.3 :	Attributs communs aux éléments MLR	100
4.4 :	Mapping canonique et lexical	101
4.5 :	Indicateurs des caractéristiques d'un élément de métadonnée	102
4.6 :	Exemple d'un élément MLR et ses attributs	104
4.7 :	Exemple d'un vocabulaire MLR	104

## Chapitre 5

5.1 :	Le modèle abstrait du <i>Dublin Core</i>	106
5.2 :	Exemple d'identifiants linguistiquement neutre	107
5.3 :	Représentation des éléments MLR dans la machine	108
5.4 :	Les éléments MLR pour un utilisateur humain anglophone	109
5.5 :	Contextes d'utilisation du MLR	111
5.6 :	Métadonnées centralisées et distribuées	112
5.7 :	Distribution des métadonnées MLR dans le Web	113
5.8 :	Ressource(X)HTML et attributs RDF	115
5.9 :	Notice MLR vers RDFa	115
5.10 :	Conversion d'une notice MLR avec XSLT en attributs RDF	116
5.11 :	Service GRDDL du W3C	118
5.12 :	Résultat d'une transformation GRDDL	119
5.13 :	Processus de conversion d'une notice MLR en RDFa, puis en RDF/XML	119
5.14 :	Modèle distribué et centralisé des métadonnées	119
5.15 :	Le MLR et le tagging	122

## Conclusion

6.1 :	Coûts de normalisation vs coûts de conversion	127
6.2 :	Technologies XML et coûts de conversion	128
6.3 :	Web sémantique et métadonnées	129

## LISTE DES SIGLES ET ABRÉVIATIONS

BSI	British Standards Institution
DCMI	Dublin Core Metadata Initiative
DOAJ	Directory of Open Archives Journal (DOAJ)
EICTA	European Information & Communications Technology Industry Association
eRDF	embedded RDF
IETF	The Internet Engineering Task Force
ISO	International Organization for Standardization
LCSH	Library of Congress Subject Headings
MACS	Multilingual Access to Subjects
MARC	Machine Readable Cataloguing
MeSH	Medical Subject Headings
METS	Metadata Encoding and Transmission Standard
MLR	Metadata for Learning Resources
MODS	Metadata Object Description Schema
NCSA	National Center for Supercomputing Applications
NISO	National Information Standards Organization
NIST	National Institute of Standards and Technology
OAI-PMH	Open Archive Initiative Protocol for Metadata Harvesting
OCLC	Online Computer Library Center
ONIX	Online Information Exchange
PURL	Persistent Uniform Resource Locator
RAMEAU	Répertoire d'autorité-matière encyclopédique et alphabétique unifié
RASUQAM	Répertoire des accès sujet de l'UQAM
RDF	Resource Description Framework
RDFa	RDF annotated
RVM	Répertoire des vedettes-matière de l'Université Laval
SGML	Standard Generalised Markup Language
SKOS	Simple Knowledge Organization System
SRI	Système de repérage de l'information
TCP/IP	Transmission Control Protocol/Internet Protocol
UQAM	Université du Québec à Montréal
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
W3C	World Wide Web Consortium
WWW	World Wide Web (aussi abrégé « Web »)
XML	Extensible Markup Language

## LISTE DES ANNEXES

Annexe 1 : Éléments de base du *Metadata of Learning Resources*

## RÉSUMÉ

Dans le contexte aussi ouvert et évolutif que celui du Web, les ressources à décrire se sont diversifiées créant ainsi de nouveaux besoins de description auxquels un schéma générique tel que le *Dublin Core* ne pouvait répondre. De nombreux schémas de métadonnées ont été ainsi développés, le plus souvent sans aucune concertation; les concepteurs étant généralement guidés par la nécessité de répondre à des objectifs spécifiques de leurs communautés respectives. Cette prolifération des schémas est à l'origine de plusieurs problèmes, notamment celui de l'interopérabilité des éléments de métadonnées.

L'interopérabilité des métadonnées peut être considérée selon deux niveaux : syntaxique et sémantique. Le niveau syntaxique porte sur la structure des messages échangés et vise à garantir la cohérence dans la manière dont les informations échangées sont représentées. C'est ici qu'entre en jeu le langage XML qui fournit un format de données standard pour permettre l'échange de données entre systèmes hétérogènes. Le niveau sémantique porte sur la signification des messages. Son rôle est de s'assurer que les échanges qui s'effectuent conservent leur sens afin que les parties communicantes aient une compréhension commune de la signification des données qu'elles s'échangent. L'interopérabilité sémantique est actuellement considérée comme le problème le plus important qui affecte la qualité de la recherche et de l'échange d'information. La solution à ce problème d'interopérabilité des métadonnées repose sur une approche centrée sur les communautés d'intérêt. En effet, les niveaux de difficulté et les défis à relever sont d'autant plus grands que l'on s'éloigne d'une communauté spécifique donnée et les coûts pour atteindre l'interopérabilité augmentent sans que la qualité des résultats soit garantie. Par contre, au sein d'un même domaine de connaissances ou d'une même communauté, l'homogénéité si relative soit-elle, réduit le taille des obstacles à surmonter et rend l'interopérabilité sémantique tout à fait réalisable.

Plusieurs techniques ont été proposées afin d'assurer l'interopérabilité des métadonnées. Le *mapping* est la technique la plus utilisée. Néanmoins, ce processus est extrêmement coûteux surtout lorsque le nombre de schémas devient important et il est peu efficace pour résoudre des problèmes de sémantique puisque, dans la plupart des cas, une intervention humaine est requise. La véritable solution au problème d'interopérabilité des métadonnées s'appuie sur les technologies du Web sémantique et plus particulièrement sur RDF et les ontologies. Malheureusement, le déploiement de ces technologies s'est avéré fort limité du fait de leur complexité et il faut recourir à des solutions transitoires telles que les microformats ou les

annotations RDF incorporées (RDFa). Comme les microformats n'offrent pas l'extensibilité et le niveau d'expressivité essentielle à l'interopérabilité, RDFa a été privilégié et est devenu une recommandation du W3C. Cette approche a été testée en prenant comme exemple la future norme ISO/IEC 19788, le *Metadata for Learning Resources (MLR)*. L'utilisation de RDFa permet de distribuer dans le Web sémantique des notices MLR centralisés dans un référentiel de ressources d'enseignement et d'apprentissage. RDFa permet également de répondre aux attentes du Web 2.0 pour la contribution distribuée à l'indexation des ressources en utilisant les étiquettes MLR. Cette approche permet d'ores et déjà de contribuer à la mise en œuvre du Web sémantique mais la simplicité de cette solution n'élimine pas, en définitive, la nécessité de passer un jour aux ontologies et au langage OWL lesquels offrent une solution générique d'interopérabilité des métadonnées. Le besoin d'outils faciles à utiliser reste un défi de taille pour la recherche sur le Web sémantique d'autant plus que l'indexation des ressources sera de plus en plus un processus distribué.

## INTRODUCTION

L'avènement de l'Internet dans les années 1960 marque le début d'une période de transformations profondes dans notre société du savoir. Réservé d'abord à la recherche militaire aux États-Unis, le « réseau des réseaux » fut, par la suite, adopté par un grand nombre d'universités et d'institutions de recherche du monde entier. D'abord axé sur les applications de messageries et d'échange de fichiers, l'Internet est devenu aujourd'hui une source d'information essentielle, voire incontournable. En effet, l'invention du Web au début des années 90 a modifié profondément l'Internet en le rendant plus convivial et accessible à tous, et en a fait le véhicule privilégié de gestion et de diffusion des contenus. Avec ses ressources informationnelles en constante augmentation, l'Internet a ouvert des perspectives considérables tant aux utilisateurs qu'aux producteurs et diffuseurs de l'information. Aujourd'hui, pour la majorité des internautes, la tendance est de chercher des informations et de diffuser des contenus d'abord sur le Web.

Comme toute révolution technologique, Internet suscite beaucoup d'enthousiasme mais aussi tout une série d'interrogations, de doutes et de déceptions. Dans notre société du savoir, la disponibilité de l'information au moment où on en a besoin, représente incontestablement un avantage stratégique indéniable au même titre que les ressources financières et humaines. Or, un des défis majeurs de l'internet aujourd'hui est le repérage de l'information. En effet, trouver la bonne information au moment opportun, relève souvent de la gageure. Pour diverses raisons, les internautes obtiennent des résultats souvent frustrants, en dépit de la prolifération des moteurs de recherche qui masquent l'entropie générale par le repérage de certaines ressources pertinentes (Larher & Milbeau, 1999, p. 38).



Malgré l'amélioration continue de ses algorithmes de recherche, Google est loin de satisfaire aux exigences de qualité d'un système de repérage de l'information (SRI) : le rappel, le bruit, le silence et la précision (<sup>1</sup>). D'après des observations empiriques dans plusieurs SRI, plus le rappel est élevé, plus faible est la précision (figure 0.1).

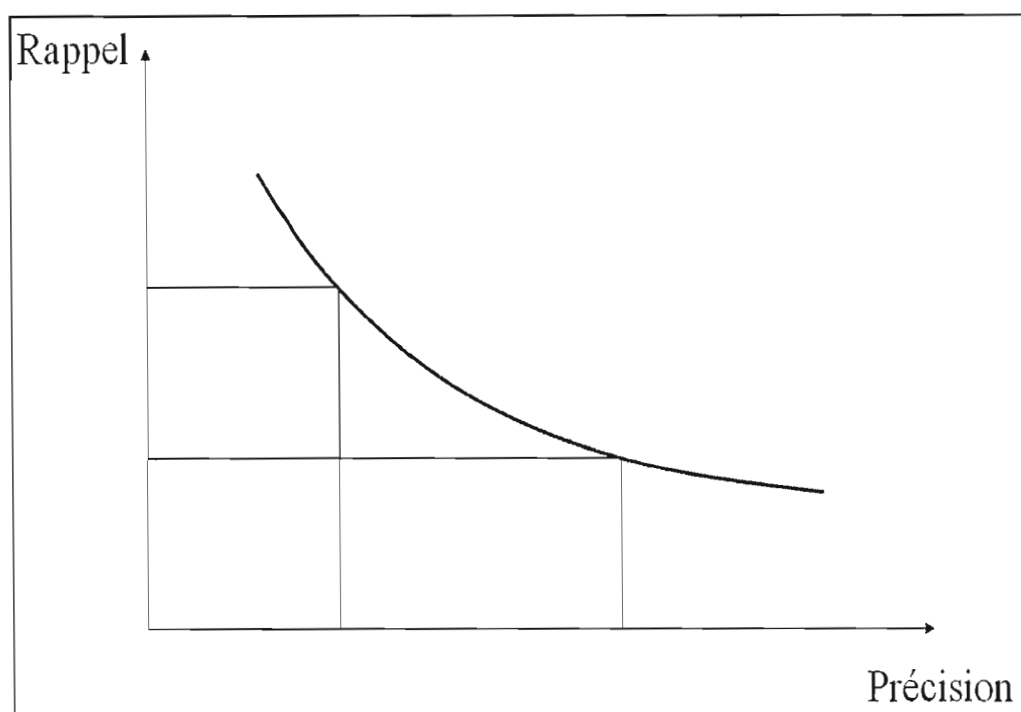


Figure 0.1 : Rapport entre le rappel et la précision (Sidhom, 2005).

Une des problématiques des moteurs de recherche comme Google est liée à la pertinence des résultats retournés, trop nombreux et présentés sous forme de liste

---

<sup>1</sup> Le *rappel* mesure le taux de documents pertinents repérés par rapport à l'ensemble des documents pertinents se trouvant dans le système. Le *bruit* mesure le taux de documents non pertinents repérés par rapport à la totalité des documents repérés. La *précision* mesure le taux de documents pertinents repérés par rapport à l'ensemble des documents repérés. Le *silence* mesure le taux de documents pertinents non repérés par rapport à la totalité des documents pertinents contenus dans le système (Sidhom, 2005).

ordonnée difficilement exploitable par l'utilisateur. De fait, une étude relativement récente montre que 81,7% des utilisateurs ne dépassent pas la troisième page des résultats (iProspect, 2004). Bref, l'efficacité relative des moteurs de recherche incite à envisager d'autres solutions. Fort heureusement, certaines de ces solutions existent depuis longtemps dans le monde des bibliothèques. Dès lors, pourquoi ne pas appliquer au Web des techniques ayant fait leurs preuves en bibliothéconomie (Larher & Milbeau, 1999, p. 38). C'est ainsi que l'idée d'utiliser les métadonnées fut mise de l'avant pour, entre autres, faciliter le repérage des ressources électroniques du Web.

Généralement définies comme des « données sur d'autres données », les métadonnées peuvent décrire aussi bien le contenu d'une ressource que les aspects relatifs à sa gestion, sa structure ou ses conditions d'accès, etc. Sur le Web, l'ajout des balises META dans l'entête d'une page HTML pour préciser son auteur ou pour en décrire le contenu à l'aide de mots-clés, en est l'illustration la plus simple.

Malheureusement, l'utilisation des métadonnées n'a pas donné les résultats escomptés. D'une part, parce que les moteurs de recherche les ont longtemps ignorées et, d'autre part, parce que leur usage a été détourné à d'autres fins, le but visé étant moins la précision de la recherche que le repérage de la page pour répondre aux besoins de référencement <sup>(2)</sup> et de marketing. En outre, les techniques utilisées par les moteurs de recherche pour déterminer la pertinence des résultats

---

<sup>2</sup> Le référencement est l'ensemble des techniques qui permettent d'inscrire un site dans les moteurs de recherche ou dans des annuaires : c'est l'art d'avoir la présence la plus massive possible de ses pages dans les moteurs (SDTICE, 2007). Cette course à une plus grande visibilité et un meilleur positionnement dans les résultats de recherche n'est pas toujours à l'avantage de l'utilisateur. Elle est souvent source de bruit.

de recherche repose plus sur l'indice de popularité d'une page que sur le sens réel du document <sup>(3)</sup>. En dépit de ces réserves, les métadonnées présentent toujours un grand intérêt car leur utilisation offre « un potentiel d'amélioration des possibilités de localisation des ressources » (Chawk, 2004, p. 3).

L'association de métadonnées aux ressources électroniques publiées sur le Web ou conservées dans divers dépôts n'a de sens que si, non seulement des outils de recherche s'en servent, mais aussi que ces métadonnées s'appuient sur des normes reconnues permettant « de garantir la cohérence des termes utilisés pour décrire les ressources électroniques au sein d'une communauté donnée » (La Passardière & Jarraud, 2004). Le besoin de standardisation des métadonnées s'est fait sentir assez tôt. En effet, pour que les métadonnées soient utilisables dans différents contextes et exploitables par différentes communautés d'intérêt, elles doivent s'inscrire dans des modèles reconnus (Chawk, 2004, p. 2). D'où l'intérêt d'avoir un schéma de métadonnées, c'est-à-dire un ensemble d'éléments dont le but est de décrire un type particulier de ressources d'information (NISO, 2004) et qui fournit généralement des instructions claires sur les valeurs de chaque élément et ses relations avec les autres.

Au milieu des années 90, un consensus international d'une cinquantaine de professionnels provenant de diverses disciplines (bibliothéconomie, muséologie, archivistique, ingénierie du Web, informatique et systèmes d'information géographique) a établi un standard de métadonnées pour Internet, appelé *Dublin*

---

<sup>3</sup> L'algorithme *PageRank* de Google ne se limite pas à afficher les pages correspondantes aux mots utilisés dans la requête mais va aussi les trier en fonction du site de provenance. La position d'une page dans la liste des résultats dépendra d'une cote évaluée en fonction du nombre de pages qui pointent sur elle et de la cote de ces pages.

*Core Metadata*. À l'initiative de l'*Online Computer Library Center* (OCLC) et du *National Center for Supercomputing Applications* (NCSA), un atelier sur invitation fut organisé du mercredi 1er mars au vendredi 3 mars 1995 à Dublin en Ohio (OCLC, 1995) dans le but de « faire une recommandation pour la description bibliographique des documents ou objets se trouvant sur le Web » (Cherhal, 2004). Le *Dublin Core*, un ensemble de quinze éléments ainsi que leurs définitions, fut spécifié en tant que dénominateur commun pour la description d'une large variété de ressources Web. Sa simplicité contraste avec la complexité du format MARC en usage dans le monde des bibliothèques depuis les années 60. De fait, dans l'esprit de ses concepteurs, le Dublin Core est d'abord destiné aux utilisateurs peu familiers avec la description des ressources (Chawk, 2004) et vise à encourager les auteurs et diffuseurs à fournir eux-mêmes les métadonnées. En effet, le format MARC n'est pas envisageable dans le contexte du Web avec ses ressources quasi illimitées (Weibel et al., 1995).

Les bonnes intentions de départ n'allaient pas tarder à se heurter à plusieurs obstacles. Tout d'abord parce que le Dublin Core se limite aux *propriétés intrinsèques* <sup>(4)</sup> de l'objet à décrire (Weibel et al., 1995) alors que d'autres éléments de métadonnées, pourtant indispensables à une meilleure accessibilité à l'information, ne sont pas pris en compte <sup>(5)</sup>. Les éléments de *Dublin Core* sont

---

<sup>4</sup> En plus de se limiter aux caractères *intrinsèques* de l'objet, le Dublin Core a adopté les principes suivants : **extensibilité, indépendance de la syntaxe, caractère optionnel, modifiable et répétable** des éléments (Weibel et al., 1995)

<sup>5</sup> Une typologie devenue classique distingue trois types de métadonnées. Les métadonnées descriptives du contenu de la ressource, les métadonnées structurelles ou techniques (caractéristiques physiques) et les métadonnées administratives (propriété intellectuelle, par exemple) (Morel-Pair, 2005, p. 3).

effectivement « gravés dans le marbre et tout élément supplémentaire appartient aux extensions » (Cherhal, 2004). En outre, l'utilisation des « raffinements » pour préciser ou étendre les éléments de base a fait qu'on est maintenant « loin de la simplicité affichée au départ » même si ce mécanisme n'empêche pas l'usage du schéma de base (Chaw, 2004). De surcroît, dans un contexte aussi ouvert et évolutif que celui du Web, la diversité des ressources à décrire et des applications disponibles a créé de nouveaux besoins de description auxquels un schéma générique tel que le Dublin Core ne pouvait répondre (Chaw, 2004).

D'autres schémas de métadonnées ont été ainsi développés, leurs concepteurs voulant d'abord répondre à des objectifs spécifiques de leurs communautés respectives. Cette prolifération de langages de métadonnées pose plusieurs problèmes, notamment celui de l'interopérabilité (Bourda & Delestre, 2005b). Et, à chaque fois qu'un nouveau schéma de métadonnées est créé, ce problème ne fait que s'amplifier. En effet, la recherche d'information dans des dépôts de ressources électroniques où cohabitent des métadonnées provenant de schémas différents se heurte à de nombreux obstacles et donne des résultats de recherche souvent imprévisibles ou incomplets (Chaw, 2004, p. 3). Dès lors, on comprend que de plus en plus, nombreux sont ceux qui recommandent la mise en place de mécanismes permettant d'assurer l'interopérabilité des éléments de métadonnées (Amerouali, 2000). Notre travail de recherche s'inscrit dans cette ligne de pensée. Quels sont les facteurs qui sont à la base du problème de l'interopérabilité des métadonnées ? Pourquoi ce problème est-il si préoccupant de nos jours et quelles sont les solutions envisagées pour y faire face ? Voilà les questions auxquelles nous tenterons de répondre dans le cadre de ce mémoire de Maîtrise en informatique.

Le présent mémoire est composé de deux parties. La première partie regroupe les trois premiers chapitres et traite des pistes de solutions au problème d'interopérabilité des métadonnées. Le premier chapitre est consacré à la problématique de l'interopérabilité des éléments de métadonnées. Les niveaux d'interopérabilité, les principaux facteurs qui sont à la base de ce problème et ses conséquences sur le repérage de l'information seront expliqués. Le deuxième chapitre est une revue de littérature faisant état des diverses solutions à l'interopérabilité des métadonnées et qui, essentiellement, proviennent du domaine de la bibliothéconomie. Le troisième chapitre se penche sur les solutions du futur et qui sont basées sur les technologies du Web sémantique. Le déploiement de ces technologies s'étant avérée fort complexe jusqu'à l'heure actuelle, la deuxième partie propose une approche pragmatique permettant de contribuer à la mise en œuvre du Web sémantique. Cette approche est basée sur RDFa, une nouvelle recommandation du W3C inspirée des microformats. En prenant comme exemple la future norme ISO/IEC 19788, le *Metadata for Learning Resources (MLR)* présenté au chapitre 4, le chapitre 5 décrit une solution permettant de distribuer les éléments de métadonnées MLR et de les rendre accessibles aux applications du Web sémantique. Nous explorons aussi les façons dont le MLR peut être utilisé dans le cadre de l'indexation distribuée des ressources afin de répondre aux attentes du Web 2.0. La dernière partie de notre travail (chapitre 6) dresse les limites et les conclusions de notre étude ainsi que des pistes de recherche futures.

## **Première partie**

**Interopérabilité des éléments de  
métadonnées : éléments de solution**

## CHAPITRE I

### PROBLÉMATIQUE

*L'interopérabilité des éléments de métadonnées* est un sujet qui fait l'objet de nombreuses recherches à l'heure actuelle. Nous verrons, dans ce chapitre, les raisons qui sont à la base de cet engouement. Mais avant tout, une clarification des concepts de base nous paraît nécessaire.

#### **1.1. Définition des concepts de base**

Commençons d'abord par «éléments de métadonnées». D'après E. Cherhal (2004), ce terme provient du vocabulaire utilisé par Dublin Core. Étant donné le contexte de sa création, celui du SGML, le Dublin Core utilise les termes « éléments » ou «attributs » là où le format MARC parle de zones ou sous-zones et où d'autres encore pourraient parler de «champs». Un élément de métadonnée est un terme qui est utilisé pour décrire une propriété ou un attribut d'une ressource (Woodley, 2005). D'un point de vue fonctionnel, il s'agit d'une «information structurée qui décrit, explique, localise une ressource et en facilite la recherche, l'usage et la gestion » (NISO, 2004). Chaque élément de métadonnée est un couple « attribut/valeur ». À chaque attribut est associé un *identifiant* qui le distingue des autres et qui peut être un nom ou n'importe quelle chaîne de caractères alphabétiques ou numériques.

Le terme « interopérabilité » est un concept clé en informatique. Les progrès de l'informatique moderne reposent d'ailleurs, en grande partie, sur cette notion qui



est omniprésente dans différents domaines de l'informatique, y compris celui des métadonnées où il joue un rôle crucial. En effet, à côté des exigences de simplicité, de modularité, de réutilisabilité et d'extensibilité, l'interopérabilité est au centre des préoccupations des différents projets de métadonnées. Plusieurs définitions de ce concept ont été proposées. Le *Livre blanc sur l'interopérabilité* de l' EICTA définit l'interopérabilité comme « la capacité de deux ou plusieurs réseaux, systèmes, périphériques, applications, composants à échanger des informations et à utiliser les informations ainsi échangées » (EICTA, 2004). Cette définition fait référence à la capacité d'échanger des données tout en minimisant la perte d'information (Morel-Pair, 2005, p. 5).

La définition du réseau INTEROP, un organisme créé en 2003 par la Commission européenne, est également intéressante surtout du point de vue de l'utilisateur. Selon cette définition, l'interopérabilité est la capacité d'un système ou un produit de travailler avec d'autres systèmes ou produits sans demander un effort spécifique à l'utilisateur (INTEROP, 2003). Dans le contexte des métadonnées, l'interopérabilité est considérée comme la capacité d'intégrer des métadonnées provenant de vocabulaires différents et de réutiliser de façon efficace ces métadonnées quelle que soit leur source (Moen, 2004). Elle implique la capacité pour différents systèmes d'échanger des informations relatives aux ressources. Ainsi, les métadonnées créées dans un système et transférées dans un autre seront traitées par ce dernier d'une manière consistante avec l'intention du créateur des métadonnées (Nilsson, 2008, p. 6).

Ces quelques définitions montrent à quel point l'interopérabilité recouvre plus que de simples questions techniques de bas niveau qui permettent aux informations de circuler correctement entre systèmes et qui sont régies par les normes et protocoles

de communication. La figure 1.1 permet de distinguer trois niveaux différents d'interopérabilité.

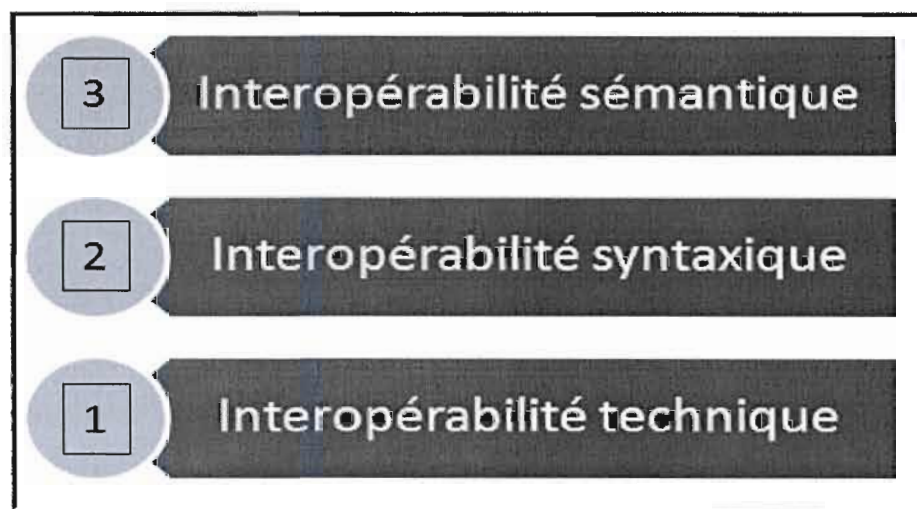


Figure 1.1 : Niveaux d'interopérabilité des métadonnées, adaptée de A. Tolk (2006)

## 1.2. Les différents niveaux d'interopérabilité

Le premier niveau fait appel aux infrastructures, protocoles et services qui permettent, entre autres, la communication des données entre ordinateurs. On peut citer, à titre d'exemple, les protocoles TCP/IP, HTTP, etc. Le détail des mécanismes mis en œuvre à ce niveau appartient au domaine des réseaux informatiques et est en dehors du cadre de notre travail. Le deuxième niveau porte sur la syntaxe, c'est-à-dire la structure des messages échangés et vise à garantir la cohérence dans la manière dont les informations échangées sont représentées. C'est ici qu'entre en jeu le langage XML qui fournit un format de données standard pour permettre l'échange de données entre systèmes hétérogènes. Ainsi, la date « **2008-02-10** » pourra être interprétée sans ambiguïté comme étant le 10 février 2008 ou le 2 octobre 2008, en fonction du schéma XML sous-jacent. De plus, Les données

exprimées en XML présentent l'avantage d'être manipulables avec les outils génériques tel que XPATH et XSLT.

L'interopérabilité syntaxique, si importante soit-elle, est cependant insuffisante. En effet, «la cohérence formelle des messages ne garantit pas, par elle-même, la cohérence des significations perçues par les différents utilisateurs d'un système » (Degoulet et al., 1997, p. 3). Un autre niveau d'interopérabilité, basé sur la signification des messages est donc nécessaire. C'est l'interopérabilité sémantique dont le rôle est «de s'assurer que les échanges qui s'effectuent conservent leur sens, c'est-à-dire que les parties communicantes ont une compréhension commune de la signification des données qu'elles s'échangent» (Heiler, 1995). Elle implique la capacité d'un utilisateur à comprendre l'information obtenue afin de pouvoir l'utiliser adéquatement (Hudon, 2001). Ainsi, lors d'un échange de données, on saura si le mot « **langue** » fait référence à un organe du corps humain ou s'il signifie un outil d'expression propre à une communauté linguistique qui l'utilise. On saura également si le mot « **baccalauréat** », selon le contexte, est un grade universitaire ou un diplôme sanctionnant la fin des études secondaires. Ces quelques exemples montrent que le problème de l'interopérabilité des métadonnées est le résultat d'une combinaison de plusieurs facteurs.

### **1.3. Les facteurs d'interopérabilité des métadonnées**

Dans son article, W.E. Moen (2004) a tenté d'élucider les différents facteurs pouvant favoriser ou non l'interopérabilité des métadonnées : multiplicité des systèmes de repérage de l'information, multiplicité des protocoles ; diversité des schémas de métadonnées, multiplicité des langues, diversité des vocabulaires et ontologies

développés pour diverses disciplines, diversité des jeux de caractères, hétérogénéité des formats de représentation des métadonnées (figure 1.2).

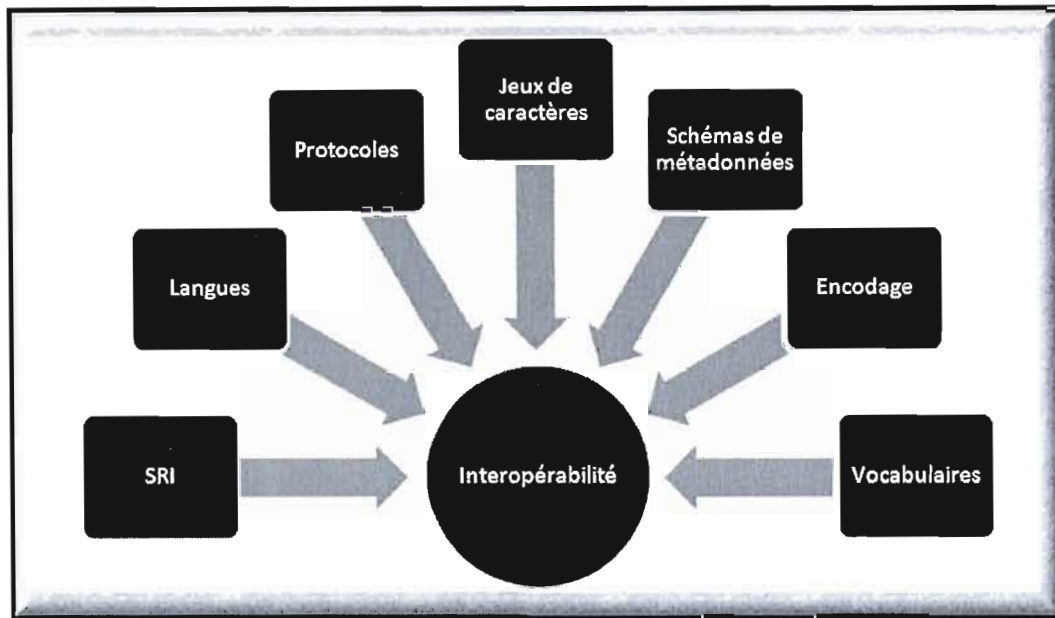


Figure 1.2 : Facteurs d'interopérabilité des métadonnées (Moen, 2004)

À ces sept facteurs s'ajoutent deux autres, identifiés par Nilsson (2008) : la diversité des modèles abstraits sur lesquels reposent les schémas de métadonnées ainsi que l'identification des éléments de métadonnées. À la lumière de son analyse, Moen conclut qu'une bonne compréhension de ces facteurs est un préalable à l'élaboration de solutions efficaces au problème de l'interopérabilité des métadonnées puisqu'une telle démarche permet de prendre les meilleures décisions quant aux actions à envisager. Il nous semble donc important d'examiner de plus près ces différents facteurs, en les illustrant, le cas échéant, par des exemples concrets.

### 1.3.1. Multiplicité des SRI

La multiplicité des systèmes de repérage d'information est illustrée de façon éloquente par la figure 1.3. Elle s'explique, entre autres, par la nécessité de répondre à des besoins d'information de différentes communautés d'intérêt. La figure 1.3 reflète la réalité du milieu universitaire mais elle est facilement généralisable à d'autres contextes. Dans ce cas-ci, différents systèmes d'information tels que les catalogues de bibliothèques, les bases de données bibliographiques, les dépôts institutionnels, etc. sont mis à la disposition des membres d'une communauté universitaire.

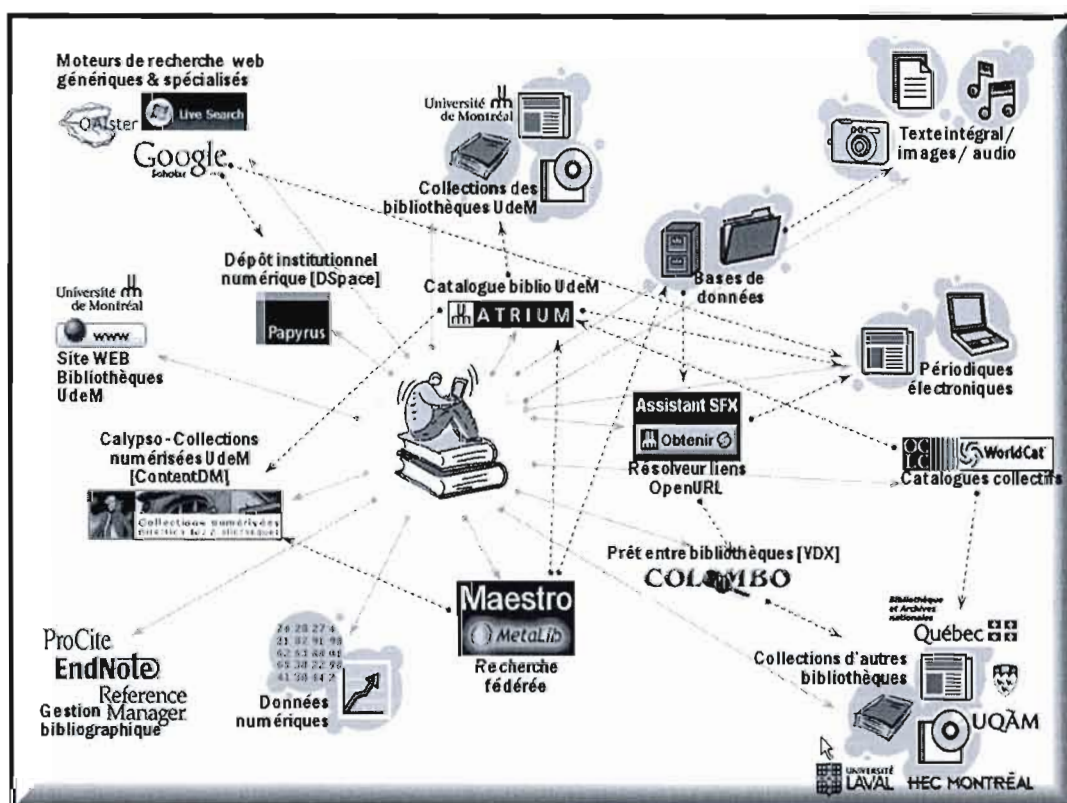


Figure 1.3 : Multiplicité des systèmes de repérage de l'information (Vézina, 2007).

Pour l'utilisateur, se retrouver dans un tel « chaos informationnel » n'est pas chose facile. L'accès à de nombreux dépôts d'archives ouvertes tel que le *Directory of Open Archives Journal* (DOAJ) <sup>(9)</sup>, *Eprints* <sup>(10)</sup>, Archipel <sup>(11)</sup> ne fait qu'amplifier le problème. Le concept relativement récent *d'obésité informationnelle* prend tout son sens. L'idéal serait qu'il puisse bénéficier d'un accès simultané et transparent à plusieurs systèmes pour effectuer ses recherches en ayant l'impression d'utiliser un système homogène et centralisé. Pour répondre à cette préoccupation et dans le but de faciliter le travail de recherche d'information dans un environnement réseau, plusieurs outils de recherche fédérée ont vu le jour. À titre d'exemple, mentionnons la possibilité de faire une recherche dans plusieurs catalogues de bibliothèques en même temps. Cependant, la qualité des résultats d'une recherche fédérée dans plusieurs SRI est tributaire, entre autres, du niveau d'interopérabilité des métadonnées utilisées par les systèmes respectifs.

### **1.3.2. Multiplicité des protocoles**

La recherche simultanée dans plusieurs catalogues de bibliothèques est rendue possible grâce au protocole Z39.50. C'est également dans cette volonté de mettre en place un cadre général pour la fédération de contenus du Web que le protocole OAI-PMH (*Open Archive Initiative Protocol for Metadata Harvesting*) a été élaboré dans le but de collecter ou moissonner les métadonnées de divers dépôts de ressources afin d'en faciliter l'accès et la visibilité. Ce protocole est supporté par

---

<sup>9</sup> [www.doaj.org](http://www.doaj.org). Au 30 mai 2009, le DOAJ donne accès à 4 184 périodiques électroniques gratuits en divers domaines.

<sup>10</sup> [www.eprints.org](http://www.eprints.org)

<sup>11</sup> Archive de publications électroniques de l'UQAM (<http://www.archipel.uqam.ca/>)

plusieurs projets de dépôts de thèses électroniques et autres publications savantes. Avec ces quelques exemples, la multiplicité des protocoles n'a pas besoin d'être démontrée. Les divers protocoles visent à rendre possible, tout en la simplifiant, la recherche dans des systèmes d'information différents. Toutefois, plusieurs obstacles doivent être surmontés et la qualité des résultats de recherche n'est pas toujours à la hauteur des attentes. La diversité des schémas de métadonnées utilisés par les différents systèmes explique en grande partie ces difficultés.

### ***1.3.3. Multiplicité des schémas de métadonnées***

La diversité des schémas des métadonnées est-elle incontournable ? Pourquoi ne pas se contenter d'un schéma de métadonnées universel ? C'est cette vision qu'avaient les concepteurs du format MARC. Si, théoriquement, l'existence d'un schéma de métadonnées universel est une solution au problème d'interopérabilité, une telle approche est plutôt utopique, surtout dans le contexte d'aujourd'hui. Il y a en effet différents besoins de description de ressources non seulement en fonction de leur nature mais aussi de leurs utilisateurs (Zen & Chan, 2006a). L'univers du Web étant accessible à tous, les métadonnées sont utilisées par diverses communautés aux intérêts parfois divergents. La généralité du modèle du Dublin Core est à la fois un atout puisqu'elle le rend applicable à tout type de ressource, mais aussi une faiblesse dès lors qu'on veut l'appliquer à un domaine précis (Ta Tuan, 2005, p. 44)

La multiplicité des schémas de métadonnées est donc inévitable. La diversité des contenus numériques, la diversité des utilisateurs et des usages font qu'il est impensable de recourir à un seul langage de métadonnées pour la description de ressources électroniques disponibles sur le Web. De nombreux schémas existent, chacun étant adapté à une application spécifique ou à un secteur d'activités tel que

celui de l'éducation. Et, même au sein d'un même domaine comme celui de l'éducation, la situation est parfois complexe du fait de l'existence des standards et des profils d'applications différents (Degoulet et al., 1997). Dans ce contexte, le partage des métadonnées n'est pas aisé. D'une part, les problèmes sont dus au fait que les noms donnés aux éléments de métadonnées sont généralement des termes tirés du langage naturel. On se heurte alors au problème de dénomination des objets car la probabilité que deux concepteurs de schémas utilisent le même nom pour représenter un même élément est très faible, généralement inférieur à 10% (Furnas et al., 1987). Le concept d'auteur pourra être représenté par différents termes tels que **<author>**, **<writer>** ou **<100>**.

D'autre part, rien ne garantit qu'un élément de métadonnée particulier, **<editeur>** par exemple, soit interprété de la même manière dans tous les contextes. C'est ici que le problème de l'interopérabilité sémantique apparaît. En effet, « il ne suffit pas de s'entendre sur la façon de nommer les choses mais aussi sur le sens qu'on accorde aux mots qu'on utilise pour ce faire » (SCTIC, 2002, p. 15). L'interopérabilité sémantique est actuellement considéré comme le problème le plus important qui affecte la qualité de la recherche et de l'échange d'information (Cousins & Sanders, 2006, p. 133).

#### ***1.3.4. Diversité des langues***

La multiplicité des schémas de métadonnées répond aussi à un autre impératif, celui de «l'adaptabilité culturelle et linguistique» qui exige une adaptation des métadonnées aux caractéristiques particulières de différentes langues et cultures d'une région, d'un pays ou d'un groupe de pays. En effet, «une infostructure sera d'autant plus utilisée que si les utilisateurs s'y retrouveront culturellement et



linguistiquement » (Hudon, 2001, p. 20). La nécessité de répondre à cet impératif se heurte à des obstacles importants. Ainsi, une recherche par sujet dans le catalogue de l'Université McGill et celui de l'UQAM rencontrera certainement des problèmes car les vedettes matières sont en anglais dans un cas et en français dans un autre.

### ***1.3.5. Diversité de vocabulaires***

La conception d'un vocabulaire pour la description des ressources exige la prise en considération des besoins spécifiques d'une communauté et le contexte socioculturel dans lequel ces ressources sont produites. Ce processus se fait généralement dans une perspective où les seuls objectifs considérés sont de satisfaire à des besoins d'une communauté donnée. Il en résulte plusieurs problèmes concernant, en particulier, la dénomination des éléments de métadonnées, leur définition et les valeurs associées à ces métadonnées. Les conséquences d'une telle situation sur le repérage de l'information sont évidentes. En effet, si la recherche dans plusieurs catalogues de bibliothèques semble donner des bons résultats grâce à l'utilisation généralisée du format MARC, du moins si l'on se limite à des critères simples tels que le titre, l'ISBN, l'auteur, il en est autrement lorsqu'on veut interroger en même temps un catalogue de bibliothèque, un catalogue de musée, un catalogue d'archives et une base de données d'articles électroniques comme celle de l'IEEE Xplore <sup>(12)</sup>. Une recherche par sujet dans ces

---

<sup>12</sup> Revues et conférences de l'IEEE (Institute of Electrical and Electronics Engineers) et de l'IET (Institution of Engineering and Technology), ainsi que les normes IEEE.

systèmes n'est pas évidente étant donné l'utilisation de vocabulaires d'indexation <sup>(13)</sup> et des schémas de métadonnées différents.

### ***1.3.6. Diversité des jeux de caractères***

À ce problème de diversité de langues, de schémas, de vocabulaires, s'ajoute celui des jeux de caractères utilisés. Les problèmes de recherche avec des expressions contenant des caractères accentués ou des caractères dits « spéciaux » est bien connu. L'adoption du standard ISO/IEC 10646 (UNICODE) devrait permettre de résoudre ces problèmes.

### ***1.3.7. Hétérogénéité des formats de représentation des métadonnées***

La multiplicité des formats d'encodage est la conséquence logique des choix qui ont été faits par les concepteurs des schémas de métadonnées qui considèrent, à tort, qu'il s'agit d'un détail d'implémentation qui n'a pas à être normalisé (Weibel et al., 1995). Ainsi, le format MARC peut être représenté sous forme d'étiquettes numériques ou de balises XML. De la même manière, le Dublin Core n'a pas un format unique de représentation ; il peut être exprimé en HTML, XML, XML/RDF en fonction de l'usage auquel il est destiné. Et d'autres types de représentations peuvent être considérés. Certes, l'hétérogénéité des formats de représentation est parfois souhaitable puisque certains formats s'apprêtent mieux que d'autres à

---

<sup>13</sup> Comme langage d'indexation matières, HEC Montréal utilise le *Répertoire des vedettes matières* de l'Université Laval ; l'UQAM a développé son propre vocabulaire, le RASUQAM alors que l'Université McGill utilise les *Library of Congress Subject Headings*.

certain types d'applications. En contrepartie, pour assurer l'interopérabilité, il faut recourir à des outils de traduction d'un format de représentation dans un autre.

La situation actuelle est caractérisée par l'existence d'une multitude de schémas ayant chacun sa syntaxe propre. On comprend alors la complexité des problèmes rencontrés lorsqu'on tente d'interroger simultanément des systèmes d'information différents, de fédérer virtuellement plusieurs dépôts de ressources électroniques ou de regrouper dans un même dépôt des ressources décrites selon des schémas différents (Chawk, 2004, p. 9). Dans un cas comme dans l'autre, on se heurte au problème d'interopérabilité.

Selon Nilsson, le problème d'interopérabilité ne se poserait pas si la création d'un schéma de métadonnées respectait les principes élaborés par Duval et ses collègues (2002a), à savoir l'extensibilité, la modularité, le raffinement d'éléments et le multilinguisme. L'interopérabilité des métadonnées irait de soi et on n'aurait pas besoin de traductions, de mapping ou de toute autre intervention humaine (Nilsson, 2008). Ce n'est malheureusement pas le cas et ce, pour deux raisons : d'une part, la diversité des modèles abstraits et, d'autre part, le mécanisme d'identification des éléments de métadonnées. Aux sept facteurs identifiés précédemment, il faut donc ajouter ces deux contraintes supplémentaires.

#### ***1.3.8. Diversité des modèles abstraits***

Une spécification de métadonnées repose, en général, sur un modèle abstrait ou métamodèle, qui spécifie les concepts utilisés et la façon de les combiner pour former une description. Les éléments de métadonnées sont définis en utilisant ce modèle. Cela permet qu'un même schéma de métadonnées, exprimé dans

différents formats, reste compréhensible d'une façon uniforme par les utilisateurs humains ou machines. L'existence des différents modèles (hiérarchique, entité-relation, structure XML) est un obstacle à l'interopérabilité des métadonnées étant donné que ces modèles diffèrent substantiellement dans la manière d'ajouter des extensions

### 1.3.9. Mécanismes d'identification des éléments de métadonnées

Un modèle abstrait en soi ne suffit pas; il faut également définir un ensemble d'éléments de métadonnées, et cela requiert l'utilisation d'un vocabulaire particulier. Nilsson (2008) distingue deux types de vocabulaires : d'un côté, le vocabulaire utilisé pour nommer les éléments et, d'un autre côté, le vocabulaire utilisé pour renseigner la valeur d'un élément de métadonnées.

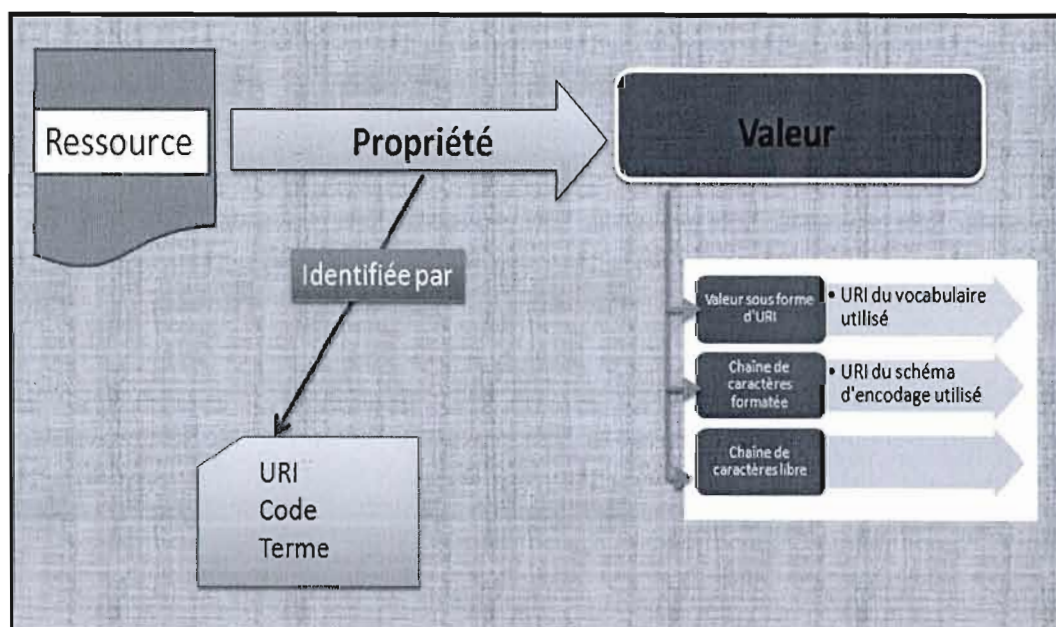


Figure 1.4 : Identification des éléments de métadonnées

Le premier est un ensemble de termes utilisés en tant qu'attributs descriptifs, tels que les quinze éléments du Dublin Core. Le deuxième est un ensemble de concepts ou de termes qui peuvent être utilisés comme valeur d'un élément de métadonnées. Nous avons déjà vu qu'un élément de métadonnées est un couple «**attribut – valeur**». Les problèmes d'identification peuvent donc se poser tant au niveau de l'attribut (ou propriété) que de la valeur de cet attribut. L'attribut peut être identifié par un URI, un code ou un terme du langage naturel. La valeur de l'attribut peut être du texte libre ou se conformer à un vocabulaire bien défini (figure 1.4). Pour décrire le sujet d'une ressource, par exemple, on peut recourir au *répertoire des vedettes-matière* de l'Université Laval, au *Library of Congress Subject Headings* ou à tout autre langage d'indexation-matière comme le thesaurus de l'éducation (EDUthès), le MeSH, etc.

L'absence d'un mécanisme universel d'identification des éléments de métadonnées et de leur valeur pose un sérieux problème d'interopérabilité. Actuellement, dans la majorité des cas, les noms attribués aux éléments de métadonnées jouent également le rôle d'identifiants pour ces attributs. Dans le cas où les modèles abstraits admettent l'utilisation des identifiants provenant du langage naturel, le multilinguisme devient un obstacle majeur. Dès lors, en comprend l'engouement en faveur des identifiants linguistiquement neutres et basés sur l'utilisation d'un code sans réelle signification. L'utilisation systématique d'un URI comme moyen d'identification des éléments de métadonnées a été fortement recommandé dès 2002 par la Commission européenne (Baker & Dekkers, 2003a) même si l'appel lancé au monde des bibliothèques à cette époque n'a pas eu l'écho espéré. Quant à l'identification de la valeur des attributs, le problème est également complexe. Les différents mécanismes utilisés offrent des niveaux différents de support au multilinguisme et d'indépendance vis-à-vis de l'application. La valeur d'un élément

de métadonnée peut être du texte libre, c'est-à-dire une chaîne de caractères compréhensible par des utilisateurs humains mais sans aucune signification pour les machines ; elle peut aussi se conformer à un vocabulaire bien défini.

Relever les défis de l'interopérabilité des métadonnées dans le contexte d'aujourd'hui n'est donc pas facile car il faut tenir compte de nombreux facteurs (linguistiques, culturels et autres). Dans un article récent, Zen et Chan (2006a) résument bien comment le problème d'interopérabilité des métadonnées se manifeste (figure 1.5) :

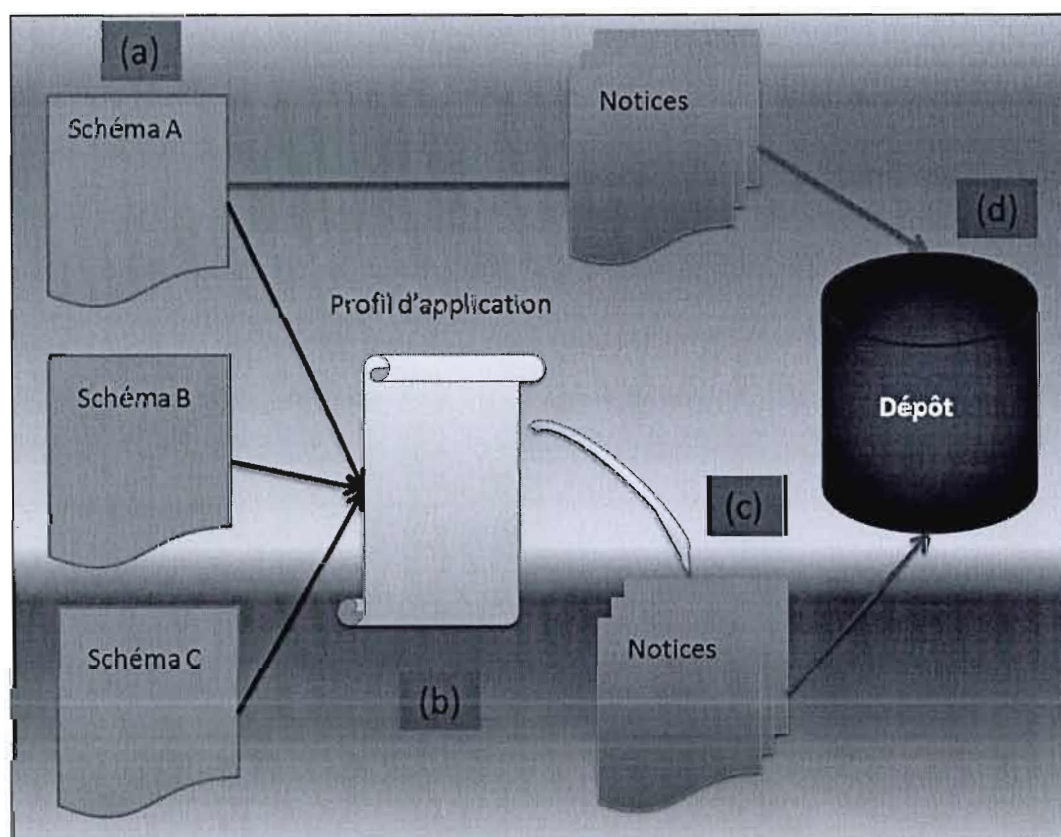


Figure 1.5: Problèmes d'interopérabilité des métadonnées (Zen & Chan, 2006b)

(a) Au départ, un schéma de métadonnées, créé pour répondre aux besoins d'une communauté donnée, est utilisé pour décrire des ressources dans le cadre d'un ou de plusieurs projets.

(b) À partir de schémas existants, un profil d'application est créé, pour répondre aux besoins d'une communauté spécifique.

(c) À l'aide de ce profil d'application, des notices de métadonnées sont créées dans un contexte particulier et conservées dans une base de données.

(d) Des notices de métadonnées appartenant à diverses collections de ressources sont « moissonnées » par un dépôt de ressources. C'est particulièrement à ce niveau que se pose le problème d'intégration des métadonnées provenant de schémas différents.

Que faire alors pour assurer l'interopérabilité des métadonnées? Si l'interopérabilité a un coût étant donné les défis à relever, l'absence d'interopérabilité n'est pas une alternative viable dans notre société du savoir. Dans le contexte du Web, la recherche d'information, pour être efficace, implique une annotation judicieuse des ressources à l'aide de métadonnées. Comme celles-ci proviennent de schémas différents compte tenu de la diversité des ressources du Web <sup>(15)</sup>, la tendance actuelle est de favoriser la réutilisation des métadonnées ne fut-ce que pour minimiser les coûts reliés non seulement au processus d'indexation des ressources mais aussi à l'élaboration de nouveaux schémas de métadonnées. L'importance des métadonnées dans la recherche de l'information n'étant plus discutable, il faut

---

<sup>15</sup> Y compris les ressources du Web invisible.



trouver des mécanismes permettant d'assurer leur interopérabilité. Le but du prochain chapitre est de faire un état de la question sur les différentes techniques actuellement utilisées pour résoudre l'épineux problème de l'interopérabilité des métadonnées. Les solutions qui font l'objet du prochain chapitre se concentrent principalement sur les métadonnées issues des schémas reconnus et qui sont fournies, soit par les professionnels de la documentation, soit par les experts du domaine.



## CHAPITRE II

### SOLUTIONS ACTUELLES AU PROBLÈME DE L'INTEROPÉRABILITÉ DES MÉTADONNÉES

Dans le chapitre précédent, nous avons passé en revue les différents facteurs qui sont à l'origine du problème d'interopérabilité des métadonnées. Nous avons également souligné à quel point il s'agit d'un problème fort complexe auquel plusieurs recherches ont été consacrées. Le but du présent chapitre est de faire une synthèse de ces travaux en faisant ressortir les principales solutions qui ont été proposées pour assurer l'interopérabilité des éléments de métadonnées et de cerner, en même temps, le travail qui reste à accomplir.

Le besoin de plus en plus grand d'un accès transparent et fiable à de systèmes d'information en réseau rend nécessaire l'interopérabilité des métadonnées. Pour résoudre ce problème, l'identification des facteurs d'interopérabilité est une étape nécessaire car elle permet de mieux intervenir sur ces facteurs (Moen, 2001). Compte tenu de la diversité des facteurs mentionnés, il importe de trouver un moyen de réduire leur impact sur l'interopérabilité. Selon E. Moen, la diversité de ces facteurs ne peut être réduite que dans le cadre d'une communauté d'intérêts donnée. Son hypothèse est que le niveau d'interopérabilité entre systèmes d'information dépend de la distance entre les communautés de pratique pour lesquelles ces systèmes ont été conçus (Moen, 2001).

Ainsi, si la recherche dans plusieurs catalogues de bibliothèques donne des résultats généralement satisfaisants, c'est parce que les données sont relativement homogènes du fait de l'utilisation d'un même schéma de métadonnées, le MARC. Par contre, la recherche simultanée dans un catalogue de musée et dans un catalogue de bibliothèque se heurte à des défis plus grands. Autrement dit, à mesure qu'on dépasse les frontières d'une communauté, la diversité des schémas et des vocabulaires utilisés augmente puisqu'il est tout à fait normal que les groupes d'intérêts différents conçoivent les choses différemment.

La mise en œuvre de l'interopérabilité ne devrait alors être considérée qu'au sein d'une communauté d'intérêt spécifique, bien délimitée et définie, ce que E. W. Moen appelle la «focal community». En effet, les niveaux de difficulté et les défis à relever sont d'autant plus grands que l'on s'éloigne d'une communauté spécifique donnée et les coûts pour atteindre l'interopérabilité augmentent sans que la qualité des résultats soit garantie. Par contre, au sein d'un même domaine de connaissances ou d'activités ou d'une même communauté, l'homogénéité si relative soit-elle, réduit le taille des obstacles à surmonter et rend l'interopérabilité sémantique tout à fait réalisable. En suivant une classification proposée par Zen (2006a) (2006b), nous verrons successivement les techniques d'interopérabilité utilisées (i) au niveau du schéma, (ii) au niveau des métadonnées, et enfin (iii) au niveau de la valeur des éléments.

## **2.1. Les méthodes d'interopérabilité au niveau du schéma**

Différentes solutions ont été proposées pour résoudre le problème d'interopérabilité en élaborant, dès le départ, un schéma de métadonnées

« compatible » avec des schémas existants. Entrent dans cette catégorie les techniques de **dérivation**, de **traduction** et d'**adaptation** de schémas existants.

### **2.1.1. La dérivation**

La dérivation consiste à créer un nouveau schéma à partir d'un schéma existant. Celui-ci, généralement plus complet ou plus complexe, sert de point de départ ou de modèle pour l'élaboration d'un schéma plus simple et plus spécialisé dans le but de mieux répondre aux besoins d'une clientèle bien déterminée. Le nouveau schéma est donc dépendant du schéma source. L'interopérabilité repose alors à la fois sur leur structure de base identique et la présence d'éléments communs. Comme exemple de dérivation, mentionnons le *Metadata Object Description Schema* (MODS)<sup>16</sup> qui est un sous-ensemble du format MARC.

### **2.1.2. La traduction**

La traduction est une technique très similaire à la dérivation. Un schéma est développé à partir des termes traduits d'un autre schéma dans une autre langue avec ou sans modification. On peut citer, par exemple, les différentes traductions de Dublin Core (<sup>17</sup>).

---

<sup>16</sup> <http://www.loc.gov/standards/mods/>

<sup>17</sup> Il existe actuellement près de 20 traductions du Dublin Core ; elles sont recensées sur le site <http://dublincore.org/resources/translations/>

### 2.1.3. L'adaptation et les profils d'application

Très proche de la dérivation est l'adaptation d'un schéma existant. Elle a pour but premier de répondre aux besoins particuliers d'une communauté donnée. On peut mentionner, par exemple, l'*Electronic Thesis and Metadata Standard* (ETD – MS<sup>18</sup>) qui est une émanation du Dublin Core auquel il emprunte treize éléments.

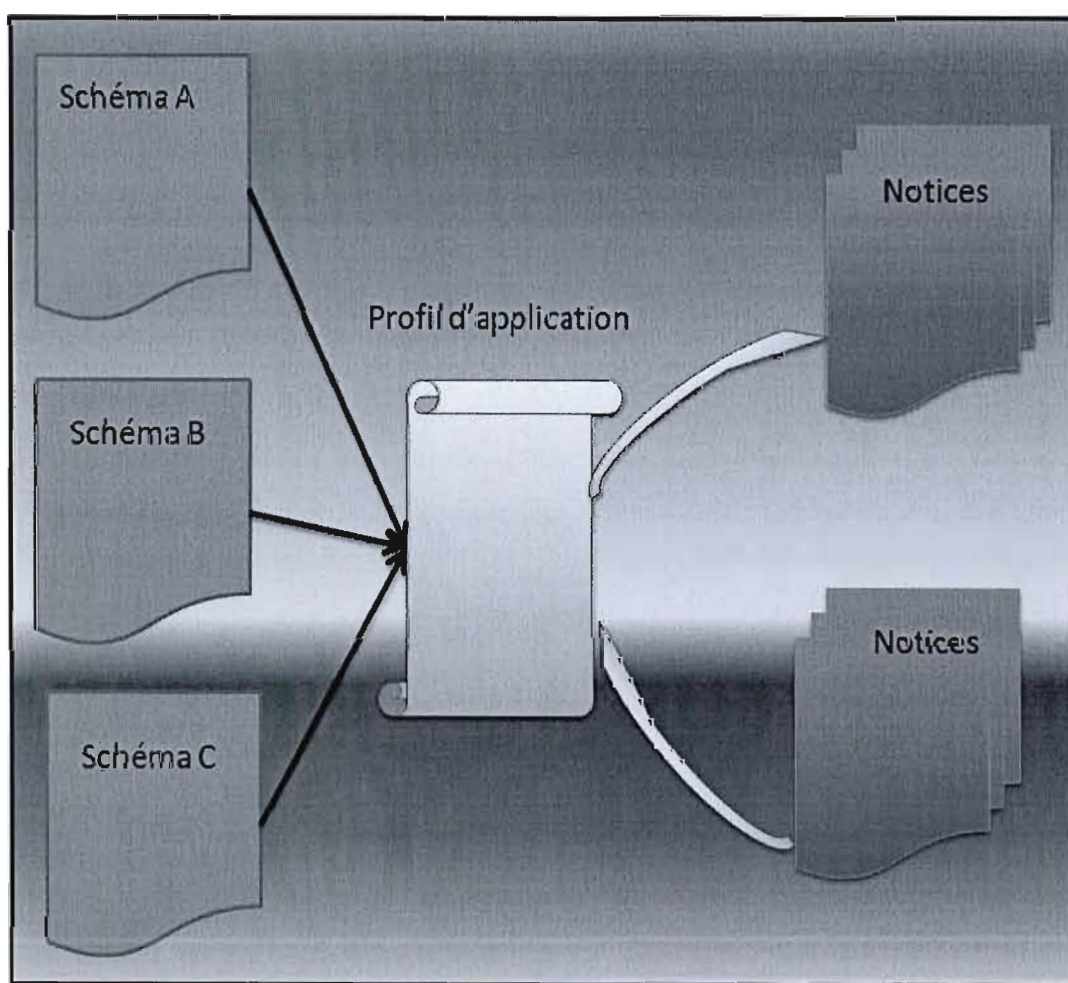


Figure 2.1 : Profil d'application (Zen & Chan, 2006a)

<sup>18</sup> <http://www.ndltd.org/standards/metadata/etd-ms-v1.00-rev2.HTML>

L'adaptation exploite le principe d'extensibilité des schémas. Ce principe permet de raffiner, d'étendre ou parfois même de simplifier les syntaxes et les sémantiques d'un schéma en fonction des environnements culturels et politiques dans lesquels une application donnée sera développée (La Passardière & Jarreaud, 2004).

En particulier, ce principe est à la base des nombreux profils d'application dont le but premier est de répondre au besoin d'une communauté d'utilisateurs. Un profil d'application (figure 2.1) est un ensemble d'éléments choisis parmi un ou plusieurs schémas de métadonnées et combinés dans un schéma composite, tout en restant interopérable avec les schémas d'origine (Duval et al., 2002b). Certains profils d'application sont constitués d'éléments issus d'un seul schéma. Mais, dans la plupart des cas, ils sont composés d'éléments issus de plusieurs espaces de noms combinés ensemble pour mieux répondre aux besoins d'une application particulière. Pour Heery & Patel (2000) et (Duval & Hodgins, 2002), dans un profil d'application, on peut (i) ajouter des métadonnées nécessaires au contexte, (ii) contraindre un élément à devenir obligatoire alors qu'il était optionnel (mais pas l'inverse), (iii) restreindre l'espace des valeurs d'un élément ou (iv) modifier les valeurs permises pour un élément. Le profil d'application permet donc d'adapter un schéma afin qu'il réponde aux besoins spécifiques et concrets d'un groupe déterminé. À titre d'exemples, on peut citer les nombreux profils d'application du LOM tel que NORMETIC au Québec, LOM-FR en France, etc. Contrairement aux schémas d'origine, les profils d'application adoptent un cadre plus contraignant car ils répondent à trois impératifs : la nécessité d'introduire un degré d'obligation des éléments (obligatoire, recommandé, facultatif), la nécessité d'ajouter de nouveaux éléments caractéristiques de l'environnement d'usage des ressources, la nécessité de définir des vocabulaires plus adaptés aux contextes institutionnels (Duval et al., 2002b).

En principe, les profils d'application devraient être compatibles les uns avec les autres tant au niveau des éléments utilisés qu'au niveau de la sémantique de ces éléments. Malheureusement, ce n'est pas le cas, essentiellement parce que ces profils sont basés sur des termes et non sur des concepts (Bourda & Delestre, 2005a). Nous reviendrons sur cet aspect plus loin dans ce chapitre.

La dérivation et les techniques affiliées exploitent les principes de modularité et d'extensibilité des métadonnées dans le but de répondre aux exigences d'une application particulière, tout en voulant préserver l'interopérabilité avec les schémas d'origine. Cependant, cette approche ne permet pas de résoudre le problème d'interopérabilité des schémas existants et totalement indépendants les uns des autres. Dans ce cas, d'autres techniques s'avèrent nécessaires. Nous décrivons ci-dessous la technique du *mapping* qui est la plus utilisée actuellement.

#### **2.1.4. Le *mapping***

Le *mapping* est la technique la plus répandue pour assurer l'interopérabilité entre différents schémas de métadonnées voire plusieurs versions d'un même schéma, par exemple lors du passage de VRA 2.0 à VRA 3.0 (<sup>19</sup>). Le mécanisme utilisé est généralement une table de correspondance qui sert à établir les équivalences entre différents éléments de métadonnées.

---

<sup>19</sup> VRA Core est un schéma de métadonnées élaboré par la Visual Resources Association pour la description des ressources patrimoniales (<http://www.vraWeb.org/>).

Lorsqu'une équivalence exacte ne peut être identifiée, on essaie de trouver une correspondance approximative en se basant sur des similarités fonctionnelles ou sémantiques (Baca et al., 2000). Les éléments d'un schéma de métadonnées sont mis en correspondance avec ceux d'autres ensembles de métadonnées qui ont une signification identique ou semblable (RCIP, 2002). Différents types d'équivalence sont possibles : une équivalence de cardinalité 1-1 lorsqu'une correspondance unique est trouvée dans le schéma de destination ; une d'équivalence multiple ou de cardinalité 1-n lorsqu'un élément est mis en correspondance avec plusieurs éléments. Dans une situation réelle, il arrive malheureusement qu'aucune correspondance satisfaisante ne soit trouvée dans le schéma cible. En outre, le *mapping* peut s'accompagner d'une perte d'information significative ou s'avérer impossible. Deux approches sont généralement utilisées dans le mapping : l'approche relative et l'approche absolue. L'approche absolue requiert une correspondance exacte entre les éléments du schéma source et ceux du schéma cible. Lorsqu'il n'y a pas d'équivalence exacte, il n'y a pas de « mapping ». Une alternative à cette approche est le mapping relatif où chaque élément du schéma source a un élément correspondant dans le schéma cible sans que l'on tienne nécessairement compte de leur équivalence sémantique ou de leur niveau de granularité. Très souvent, une intervention humaine s'avère nécessaire pour trouver une correspondance approximative. Actuellement, plusieurs schémas proposent des équivalences avec les schémas les plus populaires comme le MARC, le LOM ou le Dublin Core.

Le mapping des éléments d'un schéma avec ceux d'un autre doit tenir compte de plusieurs facteurs: la sémantique de chaque élément; le caractère obligatoire ou optionnel des éléments, la répétabilité, la relation hiérarchique entre différents éléments, les contraintes sur la valeur des éléments ainsi que le support des

éléments définis localement (St. Pierre & LaPlant, 1998). Cependant, de nombreux problèmes doivent être surmontés lors du mapping. Lorsqu'on convertit un élément, il n'y a pas toujours un équivalent exact. Dans un scénario idéal, on souhaiterait avoir un niveau d'équivalence « un à un ». En réalité, il arrive souvent qu'un terme du schéma doit être traduit par plusieurs termes du schéma cible. En outre, seuls les noms de différents éléments ainsi que leurs définitions sont pris en considération. Dès lors, la conversion des éléments d'un schéma vers un autre génère incontestablement des problèmes de qualité.

À cela s'ajoute le problème de ressources nécessaires à cette opération qui s'avère extrêmement coûteuse puisque la mise en correspondance des éléments de schémas différents requiert généralement une intervention humaine. Le problème se pose avec acuité lorsque plusieurs schémas sont mis en correspondance et que les conversions doivent être effectuées dans les deux sens. Dans le cas où le mapping porte sur un nombre important de schémas, l'identification de toutes les correspondances est loin d'être triviale. En effet, si on a  $n$  schémas de métadonnées, il faut faire  $n \times (n - 1)$  conversions pour pouvoir être totalement interopérable. Avec six schémas, il faudrait faire  $6 \times (6 - 1)$  conversions, c'est-à-dire 30 ! Ainsi, le *mapping* est un processus qui s'avère très coûteux puisque généralement, une intervention humaine est requise. Certes, des efforts significatifs ont été faits pour la mise en œuvre des outils automatisés pour le mapping de plusieurs schémas mais ils se heurtent inévitablement aux obstacles d'ordre sémantique. Et il est peu probable que dans un avenir très proche, cette opération puisse être entièrement automatisée. Ce qui explique que la plupart du temps, le mapping ne considère que les aspects syntaxiques puisqu'il est incapable de résoudre les problèmes usuels de synonymie, de polysémie et d'homonymie. Hormis cette limite d'ordre sémantique, le mapping est approprié dans le cas d'un nombre



limité de schémas. Par contre, lorsque le nombre de schémas devient important, l'utilisation d'un langage intermédiaire est conseillée. Cette technique est couramment appelée « *switching* ».

#### 2.1.5. Le « *switching* »

La technique du *switching* consiste à rendre interopérables plusieurs schémas par l'intermédiaire d'un langage de médiation. Plutôt que de faire  $n \times (n-1)$  conversions dans le cas de  $n$  schémas comme dans le cas du mapping, on utilise un schéma unique vers lequel tous les autres sont « mappés » ou mis en correspondance.

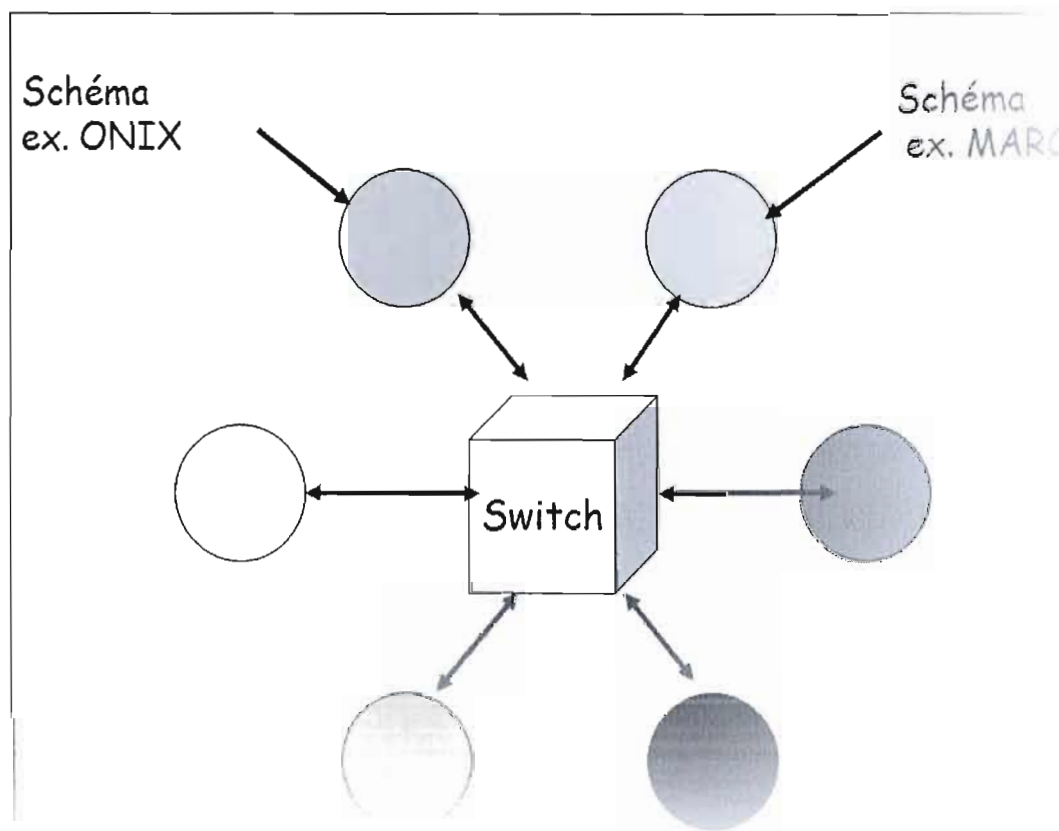


Figure 2.2 : Technique de “switching”, adapté de (Paskin, 2004)

Dans ce modèle, un schéma, existant ou non, est utilisé comme mécanisme de commutation entre des schémas différents (figure 2.2). Au lieu de faire des équivalences entre différents pairs, chaque élément du schéma est mis en correspondance uniquement avec le schéma de base. À première vue, cette approche n'est pas très différente de construire carrément un nouveau schéma avec un seul langage et ne résout donc pas totalement les problèmes d'interopérabilité. Une solution en amont consiste à définir un modèle conceptuel de référence. De tels modèles favorisent des pratiques cohérentes dans la conception des schémas de métadonnées interopérables.

#### ***2.1.6. Utilisation des modèles de référence***

Tout jeu de métadonnées représente un regard particulier appliqué au monde réel. Ce regard reflète un modèle conceptuel défini par un observateur du monde réel dans un contexte précis. L'élaboration d'un modèle conceptuel se base sur les principes de la norme ISO/IEC 11179 <sup>(20)</sup> dont le but premier est la normalisation et la gestion des éléments de métadonnées dans un registre <sup>(21)</sup> afin d'en faciliter l'échange et l'interopérabilité.

---

<sup>20</sup> ISO/IEC 11179 - *Metadata registries (MDR)*

<sup>21</sup> Un registre de métadonnées, selon le glossaire du Dublin Core, est un système qui fournit l'information sur la sémantique et la structure de chaque élément de métadonnées. Pour chaque élément, le registre donne la définition, les propriétés qui lui sont associées ainsi que la correspondance avec des équivalents dans d'autres langues ou d'autres schémas.

Cette norme comprend six parties :

*Part 1: Framework - Cadre*

*Part 2: Classification - Classification*

*Part 3: Registry metamodel and basic attributes – Métamodèle de registre et attributs de base*

*Part 4: Formulation of data definitions – Formulation des définitions de données*

*Part 5: Naming and identification principles – Principes de dénomination et d'identification*

*Part 6: Registration - Enregistrement*

La partie 1 précise le cadre d'utilisation de la norme et définit les différents concepts utilisés ainsi que la décomposition du processus de conception d'un schéma de métadonnées en plusieurs niveaux. La partie 2 détaille les méthodes de classification et d'indexation des différents éléments décrits dans la norme. La partie 3, la plus complexe, est celle qui nous intéresse particulièrement. Elle présente la structure d'un registre de métadonnées sous forme d'un modèle conceptuel. La partie 4 précise les règles à suivre pour la définition des éléments de métadonnées. La partie 5 fournit un protocole de nommage et d'identification des différents éléments d'un schéma. La dernière partie décrit les aspects organisationnels relatifs à la mise en place d'un registre de métadonnées.

L'objectif de la norme ISO/IEC 11179-3 est de définir une méthode de spécification des schémas de métadonnées. De plus, cette norme fournit une méthodologie permettant de concevoir des profils compatibles entre eux (Rebai & Labat, 2006). Bourda et Delestre, après avoir constaté que les nombreux profils d'application issus du LOM n'étaient pas interopérables, préconisent l'application de cette méthodologie lors de la conception des schémas de métadonnées et particulièrement dans l'élaboration des profils d'application. Cette méthodologie

favorise « la description standardisée des données permettant une compréhension commune de leur sémantique » (Bourda & Delestre, 2005b). La figure 2.3 représente le métamodèle de haut niveau de la norme ISO/IEC 11179-3.

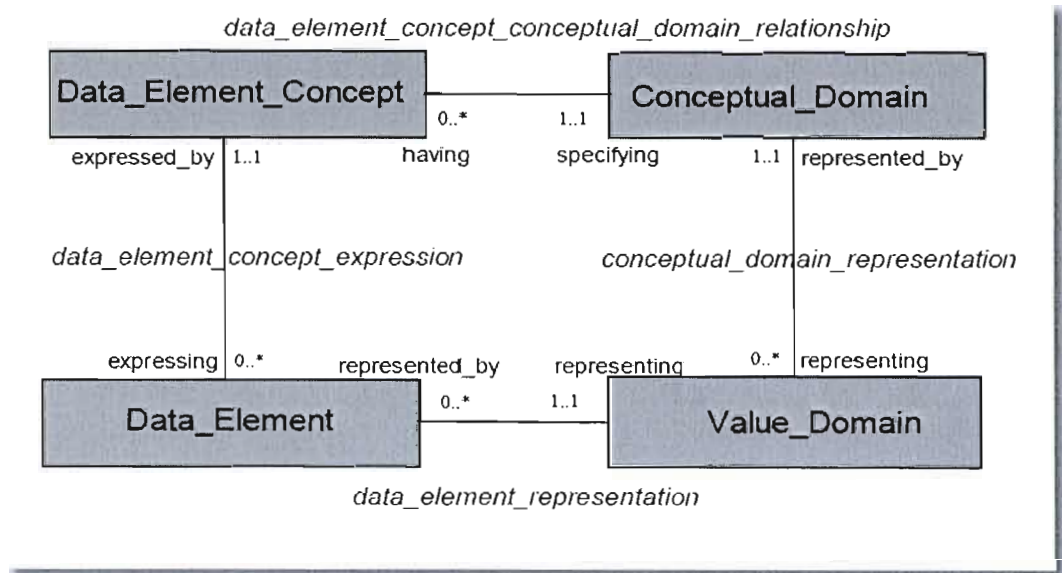


Figure 2.3 : Métamodèle de haut niveau de la norme ISO 11179-3

Selon ce métamodèle, trois composantes majeures sont au centre de la norme : les classes d'objets, les classes de propriétés et les classes de représentations (Johannis, 2005). Les classes d'objets sont des choses du monde réel qu'on peut définir explicitement et dont les propriétés suivent les mêmes règles. Les propriétés sont les caractéristiques de l'objet. La classe de représentation correspond à la forme que prendra l'élément de données. Ces composantes sont indépendantes les unes des autres et peuvent ainsi être réutilisées pour construire de nouveaux éléments de données. Les éléments de données sont représentés au moyen de domaines de valeurs. Les domaines de valeurs peuvent être non énumérés, c'est-à-dire qu'ils peuvent prendre toute valeur dans un intervalle continu. Ils peuvent aussi être

énumérés, auquel cas, ils peuvent prendre un ensemble de valeurs permises représentées par des codes, des étiquettes ou les deux. La conception d'un schéma de métadonnées se fait donc en trois étapes. La première consiste à identifier les classes d'objets. La deuxième identifie et définit les propriétés des objets. La dernière détermine le type de représentation qui sera utilisé (Johannis, 2005).

L'idée centrale de la norme ISO 11179-3 est la décomposition de la conception d'un schéma de métadonnées en plusieurs niveaux : le niveau conceptuel, le niveau des métadonnées conceptuelles, le niveau « représentation » et le niveau « classification ». Le **niveau conceptuel** (voir figure 2.4), permet de définir les entités du domaine considéré ainsi que leurs caractéristiques. Ces entités sont des instances de « Concept » et sont reliées entre elles par des instances de « Concept\_Relationship ». Les propriétés de ces entités sont, quant à elles, des instances de « Property ».

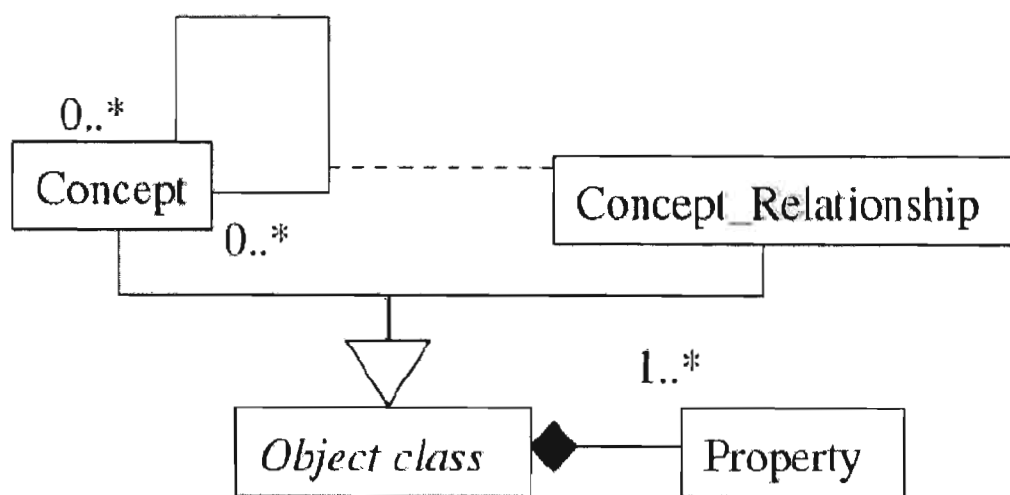


Figure 2.4 : Schéma de métadonnées : niveau conceptuel

À titre d'exemple, le concept de « Pays » peut être décrit par les caractéristiques suivantes : capitale, langue, nombre d'habitants.

Le **niveau de métadonnées conceptuel** identifie les métadonnées permettant de décrire les entités du domaine ainsi que les valeurs conceptuelles que ces métadonnées peuvent prendre.

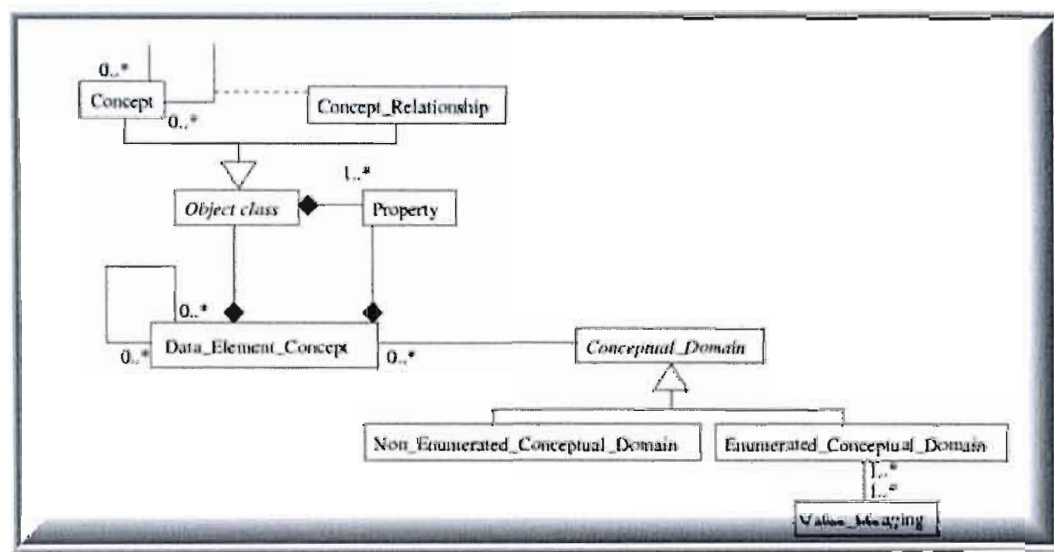


Figure 2.5 : Schémas de métadonnées : niveau de métadonnées conceptuel

La figure 2.5 montre qu'à chaque métadonnée conceptuelle (Data Element Concept) correspond un domaine de valeurs conceptuelles (conceptual domain) énuméré ou non. Ainsi, les domaines conceptuels énumérés sont associés aux métadonnées conceptuels « Identifiant de pays », « capitale du pays », « langues du pays ». Par contre, la métadonnée « nombre d'habitants » est associée à un domaine conceptuel *non énuméré* et de type *entier*.

Le **niveau « représentation »** définit, pour un contexte particulier, comment les métadonnées conceptuelles sont représentées. À ce niveau, on explicite les valeurs que peuvent prendre chaque élément de métadonnées (Data Element). À titre d'illustration, les auteurs citent la norme ISO 3166 <sup>(22)</sup> constitué de plusieurs représentations du modèle conceptuel « Pays ». En effet, dans cette norme, les noms de pays sont identifiés par leurs noms (version courte ou étendue) en français et anglais et selon leurs codes (deux ou trois caractères).

La méthodologie proposée par la norme ISO 11179-3 permet d'assurer l'interopérabilité entre différents schémas de métadonnées. Elle permet de « dissocier l'aspect conceptuel de l'aspect représentation lors de la conception d'un schéma de métadonnées » (Rebai & Labat, 2006). Dans ce contexte, les schémas de métadonnées sont des représentations de schémas de métadonnées conceptuelles. Ainsi, pour assurer l'interopérabilité entre  $n$  schémas de métadonnées, il faudra uniquement  $2 \times n$  traductions, c'est-à-dire  $n$  traductions du schéma de métadonnées conceptuelles vers les schémas de métadonnées et  $n$  traductions inverses (Bourda & Delestre, 2005b). Ainsi, la création d'un modèle conceptuel garantit l'interopérabilité des schémas de métadonnées issus de ce modèle. Ces derniers sont en quelque sorte considérés comme des instanciations d'un seul et unique modèle conceptuel. Ce qui assure leur interopérabilité. On peut facilement établir des correspondances entre différents schémas conformes à un modèle de référence et qui présentent une structure commune. Le modèle de référence constitue donc une bonne solution de l'interopérabilité lors de la conception de nouveaux schémas. En fait, lorsqu'un modèle de référence est utilisé pour créer un schéma, il n'y a pas de difficulté pour faire communiquer ces différents schémas puisqu'ils sont

---

<sup>22</sup> ISO 3166 est une norme ISO de codage de noms de pays.

construits selon les mêmes conventions. Et, à priori, ils sont compatibles entre eux. De surcroît, Bourda et Delestre (2005b) propose un outil de production de schémas de métadonnées compatibles avec la norme ISO 11179-3.

Les méthodes que nous venons de présenter s'intéressent à l'interopérabilité des métadonnées au niveau du schéma. Elles sont mises en œuvre généralement pour prévenir les problèmes d'interopérabilité. Dans le contexte où des schémas différents, conçus selon des visions différentes, sont utilisés par diverses applications, les approches pour résoudre les problèmes d'interopérabilité doivent intervenir au niveau des instances de métadonnées, c'est-à-dire au niveau des différents dépôts qui collectent et stockent ces métadonnées pour les rendre accessibles aux utilisateurs.

## **2.2. L'interopérabilité des métadonnées au niveau des instances de métadonnées**

L'interopérabilité au niveau des fiches de métadonnées se pose lorsque différents schémas de métadonnées, conçus pour répondre aux besoins de diverses communautés d'intérêt, ont été utilisés pour décrire des ressources. Comment alors intégrer ces métadonnées dans un dépôt ou les réutiliser pour créer des nouvelles notices de métadonnées ? Deux approches sont souvent utilisées : la conversion ou l'intégration des notices.

### **2.2.1. La conversion**

Dans la conversion, le problème crucial consiste à éviter les pertes d'information. Par exemple, la conversion des notices MARC en format MODS, Dublin Core ou



MARCXML. Si la conversion du MARC vers MARCXML <sup>(23)</sup> se fait sans perte d'information et vice-versa, ce n'est pas le cas lorsqu'on convertit une notice MARC en format MODS. Lorsqu'une notice MARC contient des champs qui ne sont pas prévus dans le format MODS, les données inconnues sont perdues au moment de la conversion (figure 2.6). Des notices codées en MARC21 peuvent être traduites en MODS, mais il est impossible de les reconvertir dans le format original sans perte d'information.

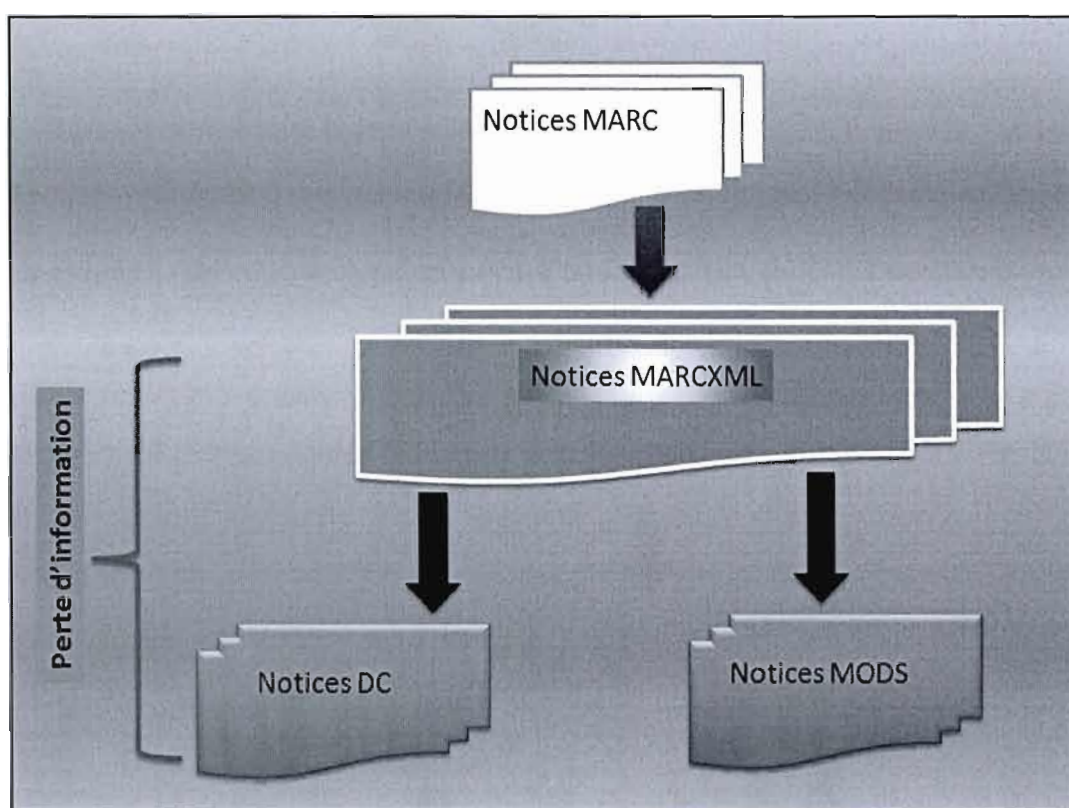


Figure 2.6: Conversion des notices issues de schémas différents

---

<sup>23</sup> MARCXML est la version XML du format MARC 21 (<http://www.loc.gov/standards/marcXML/>)

### **2.2.2. Intégration des métadonnées**

L'intégration des métadonnées repose sur le principe de réutilisation. En application du principe de modularité des métadonnées, différents éléments d'un schéma ou de plusieurs schémas peuvent être intégrés ou réutilisés pour la description d'une ressource quelconque. C'est de cette façon que les nombreuses métadonnées descriptives créées par les bibliothèques peuvent être réutilisées et intégrées avec des métadonnées techniques et administratives pour décrire de façon plus complète une ressource en ligne. Le *Metadata Encoding and Transmission Standard (METS)*<sup>24</sup> fournit un cadre permettant la réutilisation et l'intégration des éléments de métadonnées provenant de différents schémas.

METS est un schéma XML qui sert à décrire les objets numériques complexes et qui présente en un même fichier les métadonnées descriptives, administratives et structurelles reliées à ces objets. Il donne une description de la structure hiérarchique des différents objets constituant une ressource numérique, répertorie les noms et la localisation des fichiers correspondants et contient toutes les métadonnées qui leur sont associées. Le METS permet ainsi de mettre en œuvre le principe de modularité des métadonnées.

### **2.2.3. L'interopérabilité au niveau des dépôts**

Les métadonnées issues de collections différentes et rédigées à partir des schémas différents peuvent être intégrées dans un même dépôt afin que les utilisateurs puissent y avoir accès. Ce besoin s'est accentué à la suite de la Convention *de Santa*

---

<sup>24</sup> <http://www.loc.gov/standards/mets/>

Fe de 1999 sur le développement des systèmes de publication électronique et dont l'objectif était la mise en place d'un cadre général de fédération de contenus numériques, l'Open Archives Initiative (OAI). L'OAI fut à l'origine d'une approche visant à collecter les métadonnées dans un ou plusieurs services centraux en s'appuyant sur un protocole de moissonnage de métadonnées, le *Open Archive Initiative Protocol for Metadata Harvesting* (OAI – PMH)<sup>25</sup> dont la dernière version date de 2002 (<sup>26</sup>). L'idée du protocole est d'exposer les métadonnées d'une collection numérique afin d'en faciliter l'accès et la visibilité. Ce protocole définit deux types d'acteurs : les fournisseurs de données (data provider) et les fournisseurs de services (service provider). Le rôle du fournisseur de données est d'exposer les métadonnées de sa collection et de les rendre disponibles pour les autres. Les fournisseurs peuvent utiliser plusieurs schémas de description en XML mais il doit y avoir obligatoirement, dans tout serveur OAI, des métadonnées Dublin Core. L'interopérabilité repose donc sur l'utilisation du plus petit commun dénominateur, à savoir les métadonnées en format Dublin Core.

Jusqu'ici, les méthodes que nous avons analysées se concentrent particulièrement au contenant, c'est-à-dire sur le nom ou l'étiquette d'un élément de métadonnée, comme par exemple « auteur » ou « créateur » d'un document. Qu'en est-il alors de l'interopérabilité au niveau du contenu des éléments de métadonnées ? Il s'agit de l'interopérabilité sémantique et c'est particulièrement à ce niveau que les obstacles à relever sont ardu.

---

<sup>25</sup> <http://www.openarchives.org/>

<sup>26</sup> <http://www.openarchives.org/OAI/openarchivesprotocol.html>

### 2.3. Interopérabilité au niveau de la valeur des éléments

Ce problème se pose particulièrement lorsqu'on veut, par exemple, permettre aux utilisateurs d'effectuer simultanément des recherches par sujets dans plusieurs catalogues qui utilisent des langages d'indexation différents. Dans le contexte de l'Union européenne, par exemple, comment permettre à un utilisateur allemand de faire des recherches dans les catalogues des grandes bibliothèques nationales européennes en utilisant les mots clés de sa langue maternelle?

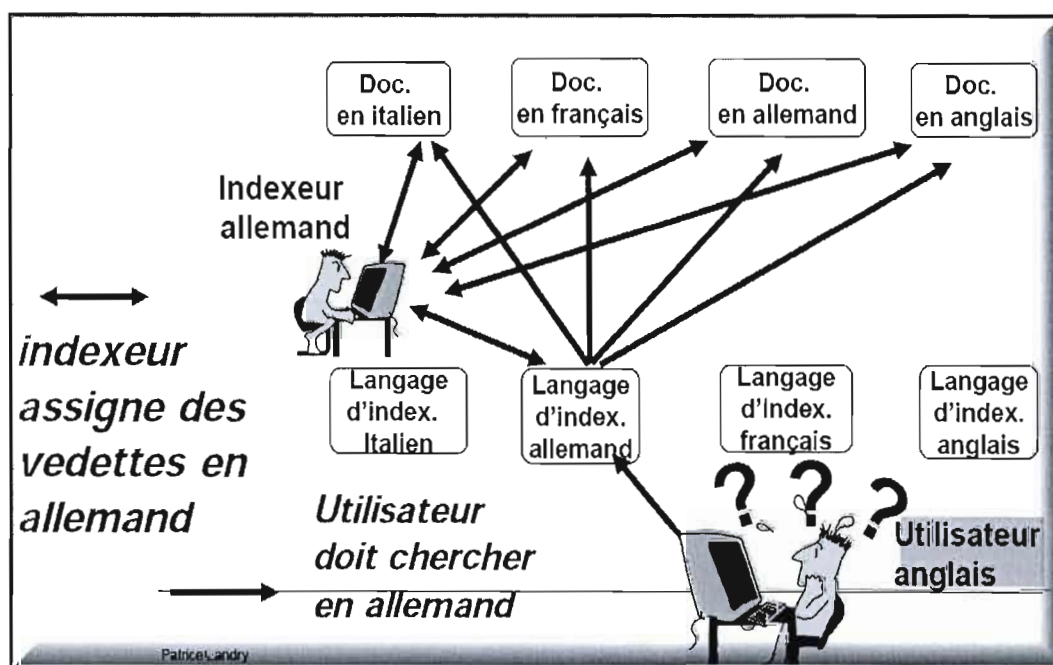


Figure 2.7 : Problème de recherche *sujet* dans un contexte multilingue (Landry et al., 2006)

Dans un environnement unilingue, l'indexeur allemand utilise les termes allemands pour représenter les sujets couverts par les documents à traiter quelle que soit leur langue de publications. Le lecteur allemand utilisera les termes allemands pour

repérer les documents pertinents dans plusieurs langues. Par contre, le lecteur anglais éprouvera des difficultés énormes lors d'une recherche *sujet* (figure 2.7). Contrairement à la recherche par auteur ou par titre, la recherche *sujet* sera confrontée à la diversité des langages d'indexation et aux différentes langues naturelles de l'utilisateur final.

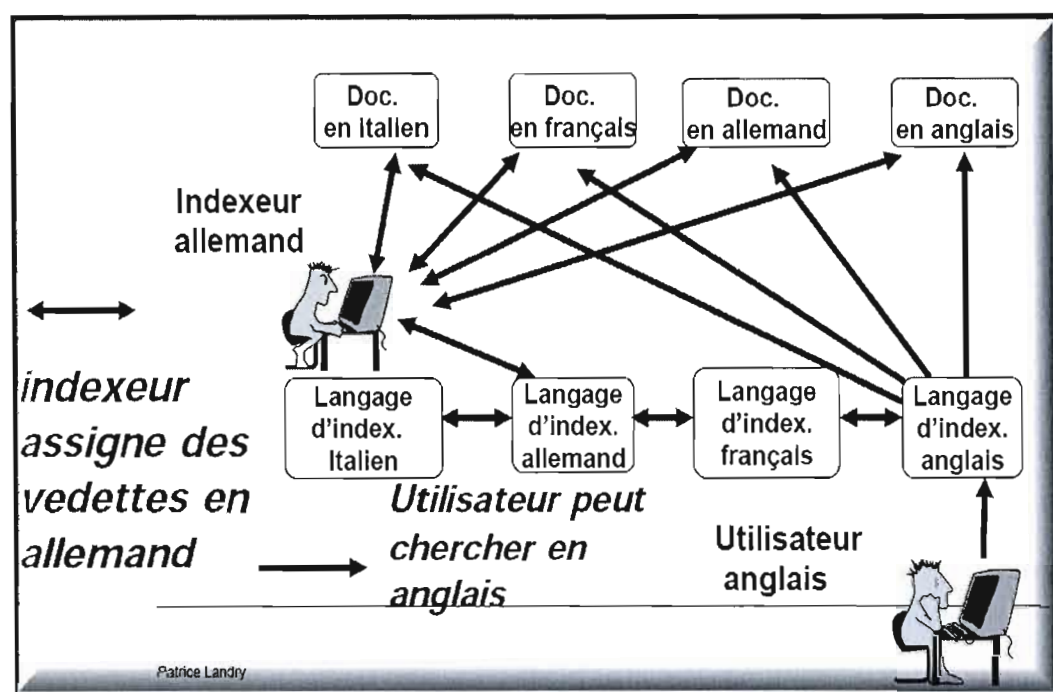


Figure 2.8: Recherche *sujet* et barrières linguistiques (Landry, 2006)

C'est pour répondre à ces préoccupations que le projet MACS (Multilingual Access to Subjects) a été mis sur pied par la Conférence des bibliothèques nationales européennes; son but étant de « trouver une solution au problème de l'accès par sujet aux catalogues en ligne dans un environnement multilingue » (Landry, 2006, p. 122). Le projet MACS vise à répondre à la question suivante : « comment faire des recherches thématiques simultanées dans plusieurs catalogues en utilisant sa propre langue » (figure 2.8)? La solution passe par l'interopérabilité des langages

documentaires utilisés. La grande question est de savoir quelle méthode utiliser pour assurer cette interopérabilité.

Un rapport publié en mai 2006 pour le compte de la Bibliothèque européenne (The European Library) a permis de «documenter les approches utilisées pour établir l'interopérabilité entre différents langages documentaires et de documenter des solutions ou approches à mettre en œuvre pour assurer l'interopérabilité entre des outils d'accès sujet au sein des dix bibliothèques nationales européennes (Landry et al., 2006). Du point de vue pratique, le but visé était de « permettre à un usager de repérer tous les documents pertinents indexés par les bibliothèques associées au projet à partir d'une seule recherche faite dans la langue de son choix : français, anglais ou allemand » (Landry, 2006).

Trois techniques furent identifiées par les auteurs du rapport : le mapping, l'utilisation d'un langage intermédiaire qui permet de naviguer parmi les termes équivalents dans plusieurs langues et l'approche fondée sur la cohabitation des vedettes sujets. La méthode du mapping fut retenue par le projet car elle respectait le principe fondamental du projet, à savoir le principe d'égalité des langues et des langages d'indexation. Elle repose sur l'établissement des liens d'équivalences entre les vedettes des différents langages d'indexation impliqués. Dans la majorité des cas, des équivalences parfaites, partielles ou complexes ont été établies, en suivant les principes préconisés par la partie 4 de la norme BS 8734 relative à l'interopérabilité des vocabulaires (BSI, 2007).

L'autre approche utilisée pour assurer l'interopérabilité est la cohabitation des éléments de métadonnées provenant de différents vocabulaires dans une même notice. À titre d'exemple, dans la notice bibliographique ci-dessous (figure 2.9), les



vedettes matières provenant du RVM et de la LCSH sont conservées dans les zones 650. L'utilisation des indicateurs donne la possibilité d'indiquer le langage utilisé : 0 pour le LCSH et 6 pour le RVM. Les mécanismes d'indexation permettront à l'utilisateur de faire une recherche *sujet* en utilisant soit la vedette française, soit la vedette anglaise.

100	1 \$aMichman, Ronald D.
596	__ \$a1
650	6\$a <b>Consommateurs riches</b> \$zÉtats-Unis
650	__o\$a <b>Luxuries</b> \$zUnited States\$xMarketing. (LCSH)
650	6\$a <b>Cibles (Marketing)</b> \$zÉtats-Unis
700	1 \$a <b>Mazze, Edward M.</b>

Figure 2.9 : Cohabitation des vedettes matières dans une même notice

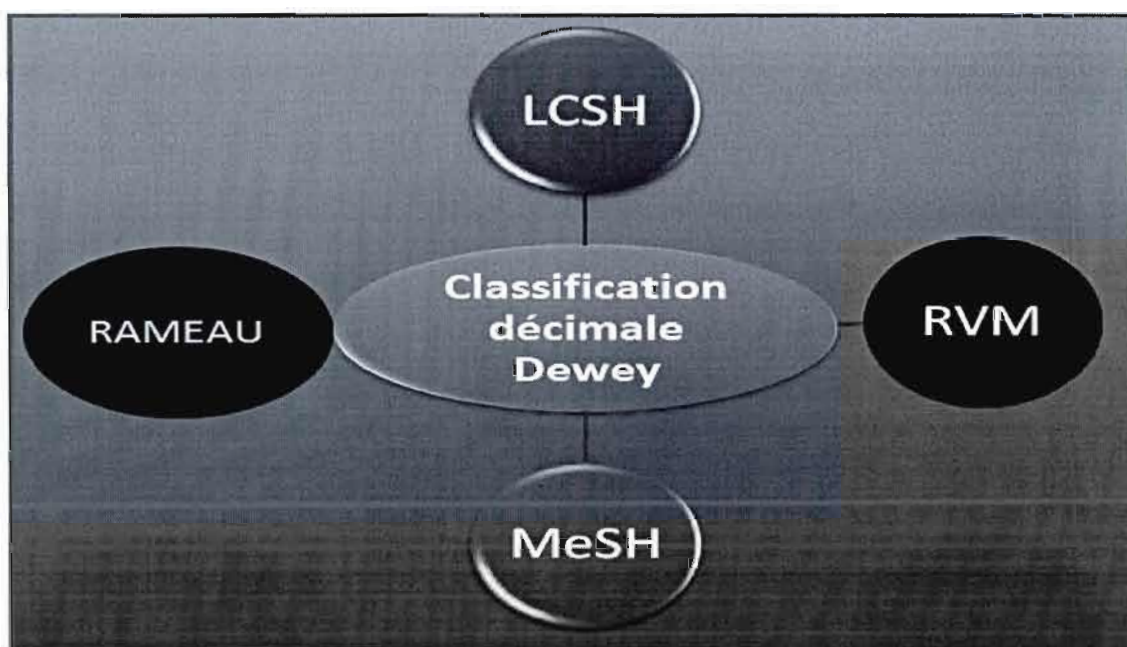


Figure 2.10 : Utilisation d'un langage intermédiaire

Une recherche sujet dans les deux langues sera ainsi possible en autant que les vedettes matières en anglais et en français auront été indexées. Comme exemple d'utilisation d'un langage intermédiaire, on peut mentionner le choix de la classification décimale de Dewey pour faire le pont entre le répertoire RAMEAU, le RVM, la LCSB et le MeSH (figure 2.10).

## **2.4. Conclusion**

Le but de ce chapitre était de broser un tableau aussi général que possible des recherches effectuées dans le domaine de l'interopérabilité des métadonnées. Il ne pouvait être question de recenser tous les travaux qui ont été publiés sur le sujet. Nous nous sommes limités aux études les plus significatives. Néanmoins, cette revue de littérature a permis de faire un inventaire des principales solutions actuellement utilisées pour assurer l'interopérabilité des métadonnées. Le mapping y occupe une place centrale. En effet, pour passer d'un jeu de métadonnées à un autre, la plupart des techniques font appel à ce mécanisme de transformation. Pourtant, le mapping se heurte à de nombreux obstacles. D'une part, l'établissement des équivalences entre plusieurs jeux de métadonnées est trop coûteux et, d'autre part, il s'agit d'un processus parfois hasardeux puisqu'il n'est pas toujours facile de traduire un élément de métadonnées vers un autre schéma surtout lorsque les pratiques diffèrent d'une communauté à une autre.

La traduction des schémas existants dans plusieurs langues est également un mécanisme fort exploité. La traduction n'étant jamais une opération neutre, la problématique des obstacles culturels et linguistiques entre ici en jeu. En effet, le travail de traduction est fortement influencé par l'appartenance du traducteur à telle ou telle culture. Dans le contexte d'aujourd'hui, caractérisé par l'attention



accordée à la pluralité des cultures et des langues, cette problématique culturelle ne peut être sous-estimée. Elle influence considérablement l'acceptation des schémas de métadonnées élaborés ailleurs indépendamment de leur utilité intrinsèque. Pour des raisons linguistiques, culturels et autres, le traducteur ne réussit toujours pas à rendre le schéma source conforme aux attentes de la clientèle cible<sup>27</sup>. La traduction ne crée pas toujours un cadre favorable à l'appréciation et à l'acceptation d'un schéma de métadonnées par les clientèles issues d'un contexte culturel différent de celui pour lequel ce schéma a été conçu à l'origine.

L'acceptabilité d'un schéma de métadonnées est fonction de l'adaptabilité culturelle et linguistique qu'il permet. Or, les schémas de métadonnées les plus connus sont conçus au départ en anglais. Ils ont fait l'objet de nombreuses traductions mais celles-ci n'ont pas toujours eu le succès escompté en dépit des efforts investis. La traduction ne garantit pas toujours l'interopérabilité sémantique puisque l'interprétation des messages se fait toujours dans un cadre de référence culturel particulier propre au traducteur.

Bref, en dépit des nombreuses recherches mentionnées dans ce chapitre, force est de constater que les solutions actuellement proposées ne sont pas exemptes de défaut. La plupart d'entre elles se focalisent principalement sur les noms des éléments de métadonnées, ce qui constitue une limite fondamentale. Et, il existe encore aujourd'hui des problèmes non résolus. En particulier, les solutions actuelles ne fournissent pas une réponse satisfaisante aux problèmes d'interopérabilité sémantique et d'adaptabilité culturelle et linguistique. En outre, les problèmes nouveaux soulevés par le phénomène « d'indexation de masse » par le **tagging** est

---

<sup>27</sup> Par exemple, on voit souvent le terme anglais *language* traduit par *langage* en français !

complètement escamoté. Le but du prochain chapitre est de montrer que la solution à ces problèmes réside dans l'utilisation des technologies du Web sémantique qui offrent « une solution générique d'interopérabilité des métadonnées » (Nogueras-Iso et al., 2004).

## CHAPITRE III

### LE WEB SÉMANTIQUE ET L'INTEROPÉRABILITÉ DES MÉTADONNÉES

Le Web sémantique est un sujet à la mode bien qu'il fut conceptualisé dès 2001 par Tim Berners-Lee dans un article publié dans la célèbre revue américaine *Scientific American* (Berners-Lee et al., 2001). Les différentes technologies (voir figure 3.1) qui le sous-tendent offrent des opportunités nouvelles pour l'interopérabilité des métadonnées (Daconta et al., 2003).

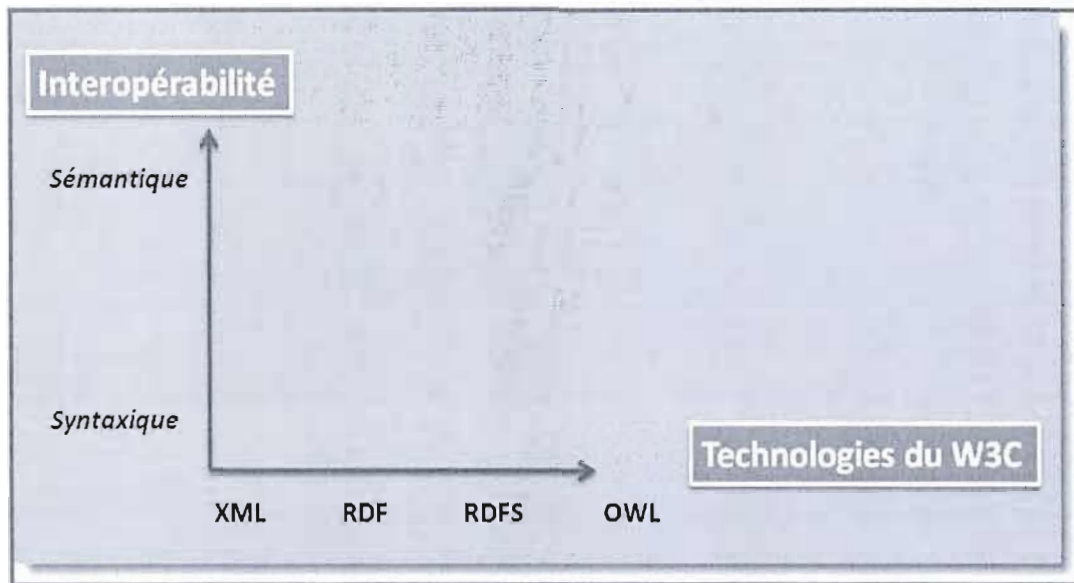


Figure 3.1 : Technologies du Web sémantique et interopérabilité (Daconta et al., 2003)

Dans ce chapitre, nous présenterons ces technologies. Nous expliciterons d'abord le concept de Web sémantique. Nous verrons ensuite les limites du Web actuel qui justifient la mise en place d'une nouvelle vision du Web. Nous décrirons par la suite les différentes technologies du Web sémantique en mettant l'accent sur l'apport de chacune d'elles à l'interopérabilité des métadonnées. Nous terminerons ce chapitre en jetant un regard particulier sur l'état actuel de l'implantation du Web sémantique.

### 3.1. Le Web sémantique

Selon Tim Berners-Lee (2001), le Web sémantique est une extension du Web actuel, dans lequel on donne à une information un sens bien défini pour permettre aux ordinateurs et aux utilisateurs humains de travailler en coopération.

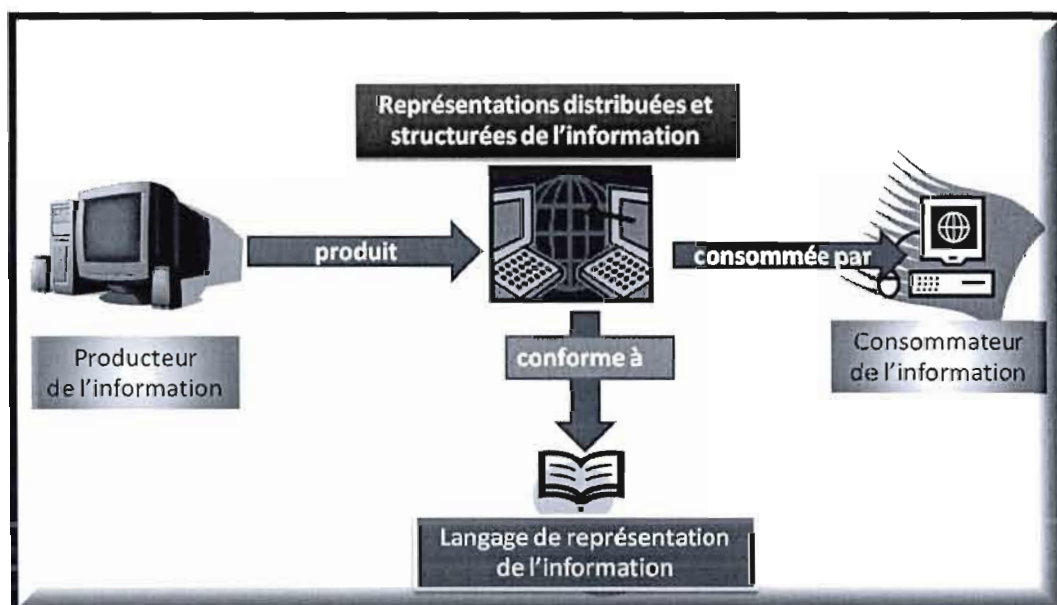


Figure 3.2 : Représentation de l'information pour le Web sémantique (Lacy, 2005, p. 14)

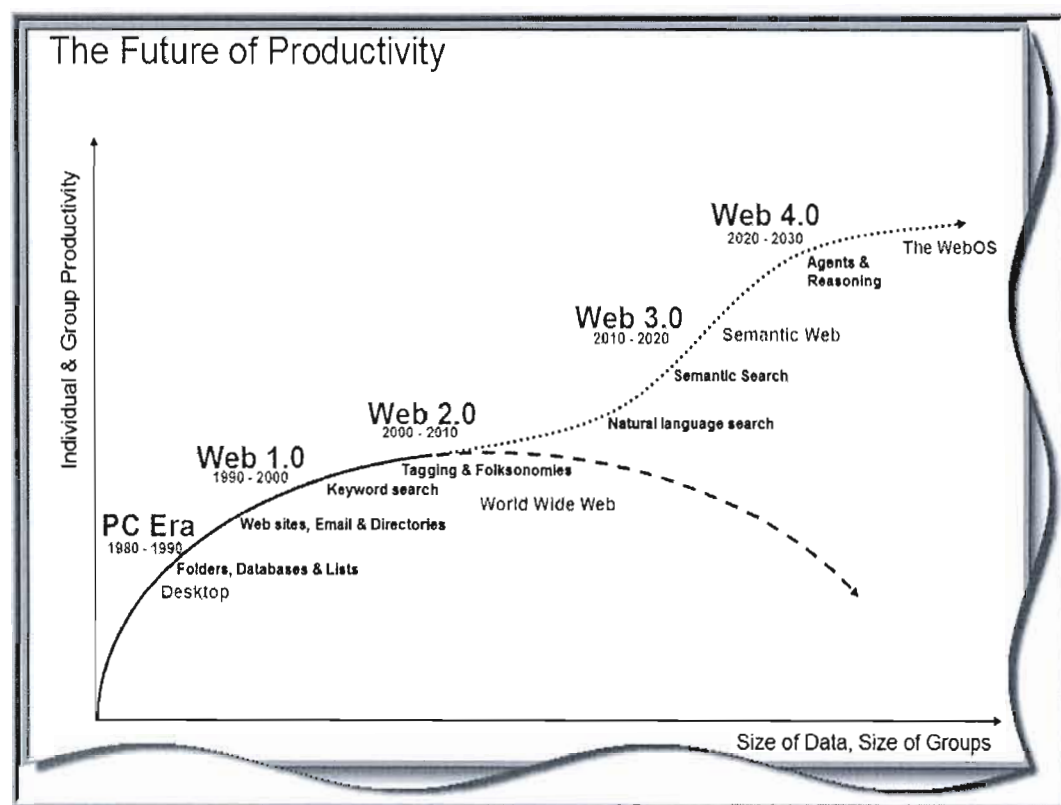
Le Web sémantique ajoute aux ressources existantes une couche de métadonnées qui les rend exploitables par les ordinateurs (Aussenac-Gilles, 2007). Les métadonnées sont alors au centre de cette vision. En effet, chaque ressource est dotée de métadonnées sémantiques définies à l'aide de langages de description de ressources (Ghafour, 2003). Ces métadonnées renferment des sémantiques non ambiguës permettant d'automatiser les traitements. Ainsi, les ressources du Web, en plus des balises HTML, doivent se doter d'annotations sémantiques dont le but premier n'est pas d'assurer leur présentation mais l'appréhension de leur contenu par divers outils logiciels (Euzenat, 2003). Contrairement au HTML qui ne donne que des liens sans sémantique, dans le Web sémantique, les liens entre différentes ressources ont un sens utilisable par les machines. Dans cette perspective, le Web sémantique n'est pas un Web de documents mais plutôt un Web de données distribuées, réutilisables et compréhensibles par la machine grâce à des langages de représentation de l'information (figure 3.2).

### **3.2. Pourquoi le Web sémantique**

Au cours de ses premières années, le W3C a développé plusieurs recommandations « centrées sur la présentation des contenus et leur proximité symbolique par les liens hypertextes » (Porquet, 2005). Ces recommandations ont fait du Web la plus grande plate-forme d'accès à l'information. Très rapidement, le Web est devenu victime de son succès. Depuis quelques années, il souffre d'un problème de maîtrise des contenus en raison même de leur volume. Pour les concepteurs du Web sémantique, le Web actuel est conçu pour être interprété par des humains car le langage naturel ne présente pas de signification pour les machines car il est non-formel (Menon, 2004). Ainsi, le Web actuel est à la fois la base et la raison d'être du

Web sémantique» (Aussenac-Gilles, 2007). Autrement dit, le Web sémantique est la réponse logique aux insuffisances de Web actuel.

Dans sa vision, Tim Berners-Lee voulait un Web sous forme d'un réseau sémantique de données accessibles aux humains et aux machines. Malheureusement, les premiers ont été privilégiés, les machines se limitant à la capacité de traiter l'information sans être capable de l'interpréter.



[http://novaspivack.typepad.com/nova\\_spivacks\\_weblog/2007/03/beyond\\_keyword\\_.html](http://novaspivack.typepad.com/nova_spivacks_weblog/2007/03/beyond_keyword_.html)

Figure 3.3 : Productivité de la recherche par mots-clés sur le Web (Spivack, 2007a)

Certes, les machines jouent un rôle prépondérant dans le Web d'aujourd'hui comme en témoignent les différents classements qui placent les moteurs de recherche

parmi les sites les plus consultés. En effet, sans les moteurs de recherche, l'expérience Web ne serait pas ce qu'elle est aujourd'hui. Toutefois, les moteurs de recherche sont incapables de trouver la réponse exacte aux questions posées et se limitent plutôt à proposer une liste plus ou moins pertinente de candidats susceptibles de fournir la réponse (Feigenbaum & Torres, 2006) <sup>(28)</sup>. De surcroît, N. Spivack (2007a) montre qu'à mesure que le Web se complexifie et prend des proportions énormes, la recherche par mots-clés devient moins efficace et semble avoir atteint son apogée car, tout comme la recherche en langage naturel, elle n'est pas capable d'interpréter la structure de l'information. Seul le Web sémantique permettra d'améliorer la productivité du processus de la recherche sur le Web (figure 3.3).

Pour relever ces défis, il faut faire un pas de plus afin que les machines puissent fournir des données agrégées. Entre autres, il faudra disposer des mécanismes permettant de définir la sémantique associée aux ressources de manière à faciliter leur exploitation automatique par l'ordinateur (Ghafour, 2003) tout en facilitant la cohabitation des utilisateurs humains et machines sur le Web, les seconds permettant aux premiers d'améliorer leur productivité dans la recherche d'information (Feigenbaum & Torres, 2006). L'objectif du Web sémantique est donc de libérer les utilisateurs humains d'une partie du travail de recherche, d'exploitation et d'intégration des résultats, grâce aux capacités accrues des

---

<sup>28</sup> Le 18 mai 2009, Stephen Wolfram, créateur du logiciel *Mathematica*, lançait **WolframAlpha** (<http://www.wolframalpha.com/>), un moteur de recherche intelligent capable de fournir des réponses à des questions factuelles posées par l'internaute au lieu de retourner une liste de liens contenant peut-être la réponse !

machines à accéder aux contenus des ressources et à effectuer des raisonnements sur ceux-ci (Aussenac-Gilles, 2007).

L'autre problématique du Web actuel est de trouver des informations fiables et crédibles. À cet égard, le Web sémantique propose « des mécanismes permettant de qualifier les connaissances afin d'augmenter le niveau de confiance des utilisateurs » (Laublet et al., 2003) et donne aux agents intelligents un rôle clé dans la recherche de l'information pour une meilleure adéquation entre les résultats de recherche et les profils de l'utilisateur.

### 3.3. Les technologies du Web sémantique et l'interopérabilité

Comme on l'a vu plus haut, le Web sémantique n'est pas une révolution et ne vise pas à remplacer le Web actuel; il en est le prolongement.

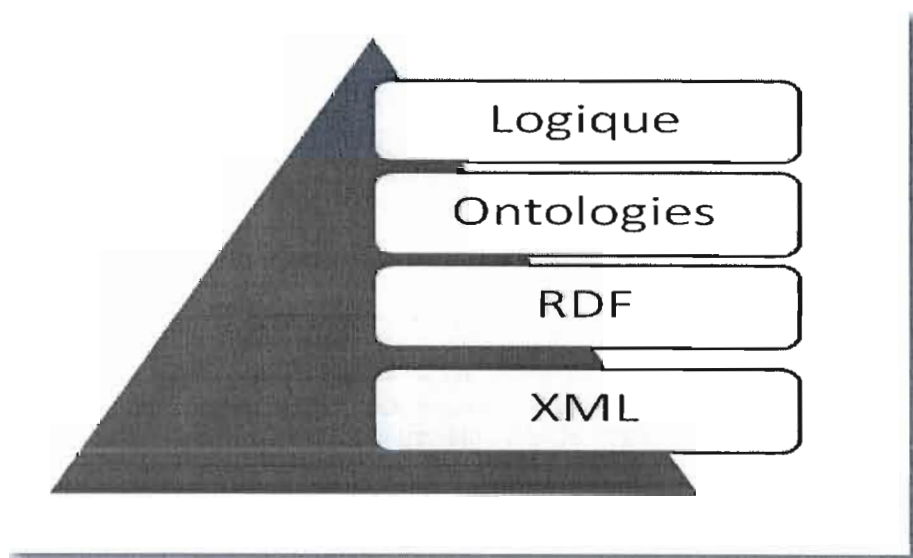


Figure 3.4 : Couches de base du Web sémantique selon Hori (2003)



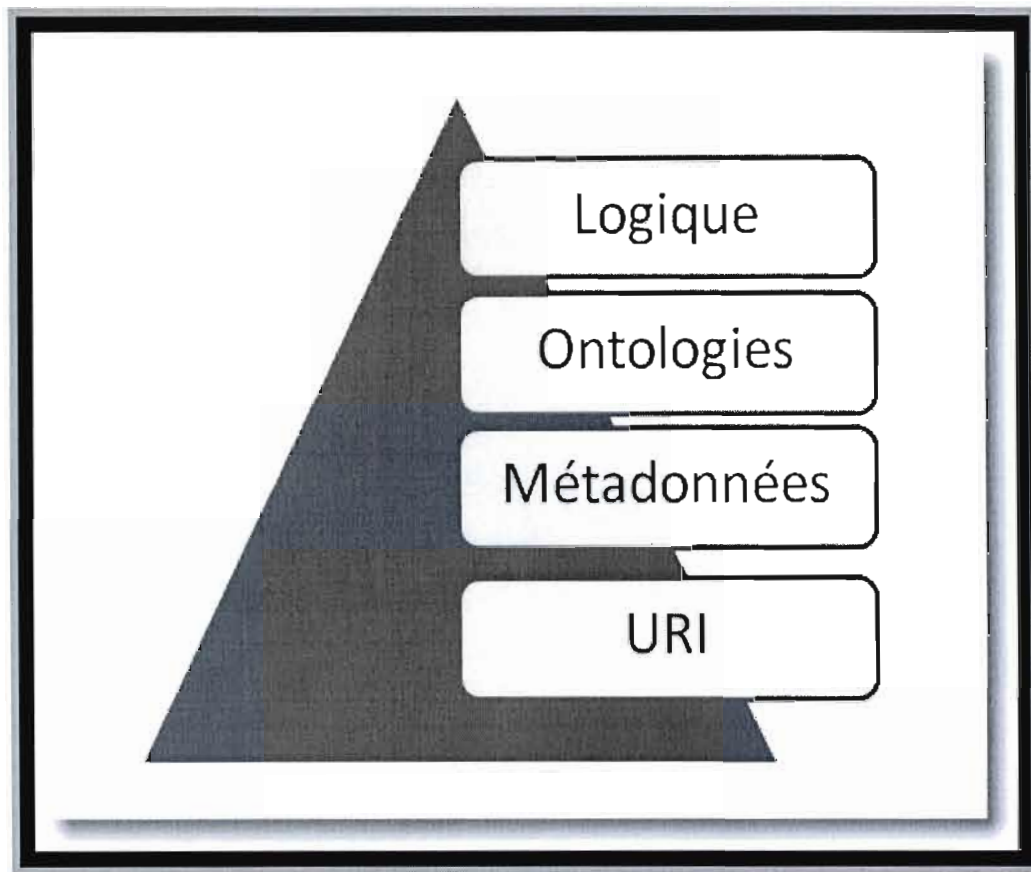


Figure 3.5 : Modèle conceptuel du Web sémantique selon Geroimenko (2006, p. 5)

Du point de vue technologique, il est souvent représenté comme une architecture en couches qui permet une approche graduelle dans le processus de normalisation et d'acceptation des ces technologies par les utilisateurs (Ta Tuan, 2005). Selon Hori (2003), cette architecture est composée de quatre couches (figure 3.4). Cette architecture est proche de celle V. Geroimenko (2006) qui distingue également quatre couches (figure 3.5).

Mais c'est l'architecture présentée par T. Bernes-Lee (figure 3.6) qui fait autorité aujourd'hui. Ambition d'une telle architecture ? Permettre aux agents de repérer l'information et de valider celle-ci, reconnaître le niveau de confiance accordé par

les utilisateurs à différentes sources d'information, en plus d'offrir un mécanisme de logique suffisamment développé pour valider les résultats et justifier les choix réalisés dans la sélection de l'information (Davis, 2007). Nous allons détailler ses différentes composantes dans la suite de cette section.

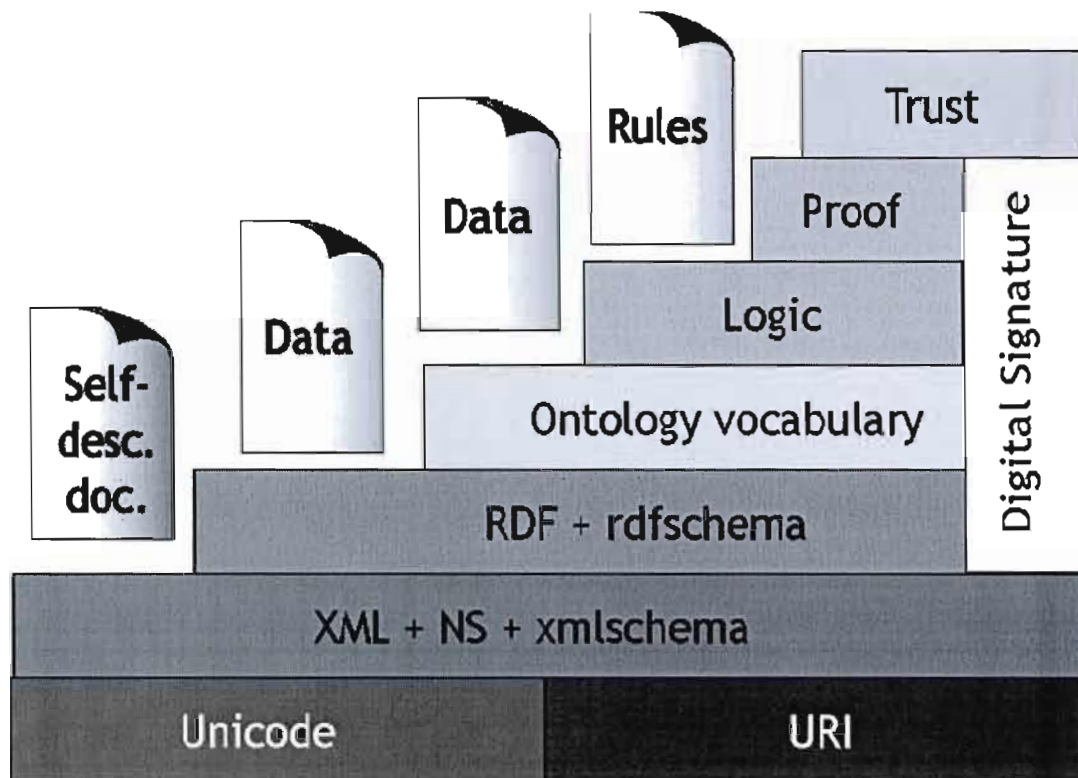


Figure 3.6 : Le Web sémantique selon T. Berners-Lee <sup>(29)</sup>.

La figure 3.6 nous montre une pile de langages ou protocoles chacun jouant un rôle particulier. Les premières couches de cette architecture sont déjà opérationnelles.

<sup>29</sup> <http://www.w3.org/2000/Talks/1206-XML2k-tbl/slide10-0.HTML>

De fait, les langages XML, RDF et OWL sont des recommandations du W3C depuis quelques années.

### **3.3.1. Couche de base**

Au niveau le plus bas, on retrouve le standard UNICODE (ISO/IEC 10646) qui a été choisi à cause de sa capacité à répondre aux besoins de différentes langues et écritures, sans oublier que les documents XML utilisent par défaut le jeu de caractères UNICODE. On retrouve aussi à ce niveau la spécification RFC3986 <sup>(30)</sup> sur les URI (Uniform Resources Identifier) ou identificateurs uniformes de ressources. Fruit d'une collaboration entre l'IETF et le W3C, cette spécification joue un rôle crucial dans le Web car elle permet d'identifier tout type de ressource qu'elle ait une existence réelle ou non. Dans le contexte des métadonnées, l'URI représente un avantage à double titre. D'une part, lorsqu'une ressource est identifiée par un URI, n'importe qui peut créer n'importe quelles métadonnées sur cette ressource. D'autre part, il est possible d'agréger dans une même notice, des métadonnées provenant des schémas différents.

Cette agrégation est rendue possible par le fait que ces métadonnées réfèrent à une même ressource. Une résolution sur les identificateurs d'éléments de métadonnées adoptée en novembre 2002 par des organismes de normalisation de métadonnées <sup>(31)</sup> confère aux URI un rôle encore plus important (Baker & Dekkers, 2003b).

---

<sup>30</sup> Uniform Resource Identifiers (URI): Generic Syntax (RFC3986). Disponible à l'adresse <http://www.ietf.org/rfc/rfc3986.txt>.

<sup>31</sup> Cette résolution, signée le 18 novembre 2002, est plus connue sous le nom de « *CORES Resolution* ». Elle a été signée, à l'initiative de l'Union européenne, par des organismes responsables

L'étude de M. Nilsson (Nilsson, 2008) montre cependant que les principes de cette résolution ne sont pas suivis systématiquement ; ce qui est à l'origine de certains problèmes d'*harmonisation* <sup>(32)</sup> de différents schémas de métadonnées.

### 3.3.2. Couche XML

La famille des langages XML constitue la deuxième couche. Dérivé du langage SGML <sup>(33)</sup> qu'il se propose de simplifier (figure 3.7), le langage XML est une recommandation du W3C <sup>(34)</sup>.

XML est au cœur du Web sémantique car il est à la base des standards développés pour répondre aux besoins de cette nouvelle génération du Web. Ce qui permet au Web sémantique d'exploiter tous les outils, langages et technologies développées autour du XML (Laublet et al., 2003) aussi bien pour la création, la transformation, la présentation des documents que la description sémantique de leur contenu (figure 3.8). Et il est fort probable que les standards qui s'ajouteront à l'architecture actuelle soient également basés sur XML (Geroimenko, 2006, p. 13).

---

de nombreux schémas de métadonnées tels que MARC, GILS, ONIX, UNIMARC, Dublin Core, IEEE/LOM, DOI, CERIF.

<sup>32</sup> Mikael Nilsson utilise le terme « *harmonisation* » plutôt qu'*interopérabilité*.

<sup>33</sup> ISO/CEI 8879 :1986 Traitement de l'information – systèmes bureautiques : langage normalisé de balisage généralisé (SGML), 1986. 164p.

<sup>34</sup> W3C XML recommendation – Extensible Markup Language (XML) 1.0. disponible à l'adresse <http://www.w3.org/TR/REC-XML>

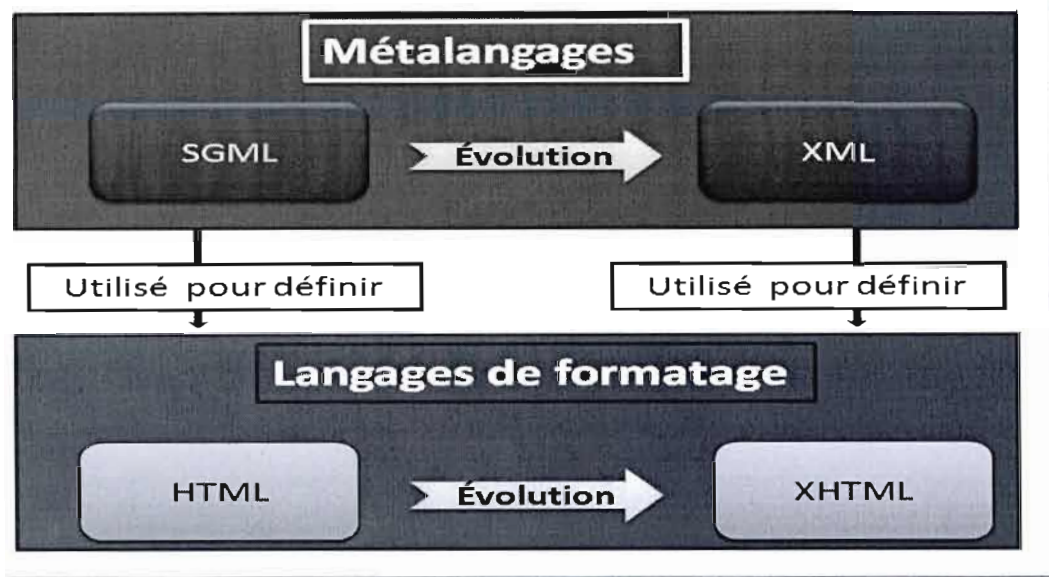


Figure 3.7 : Évolution des langages de balisage (Lacy, 2005, p. 62)

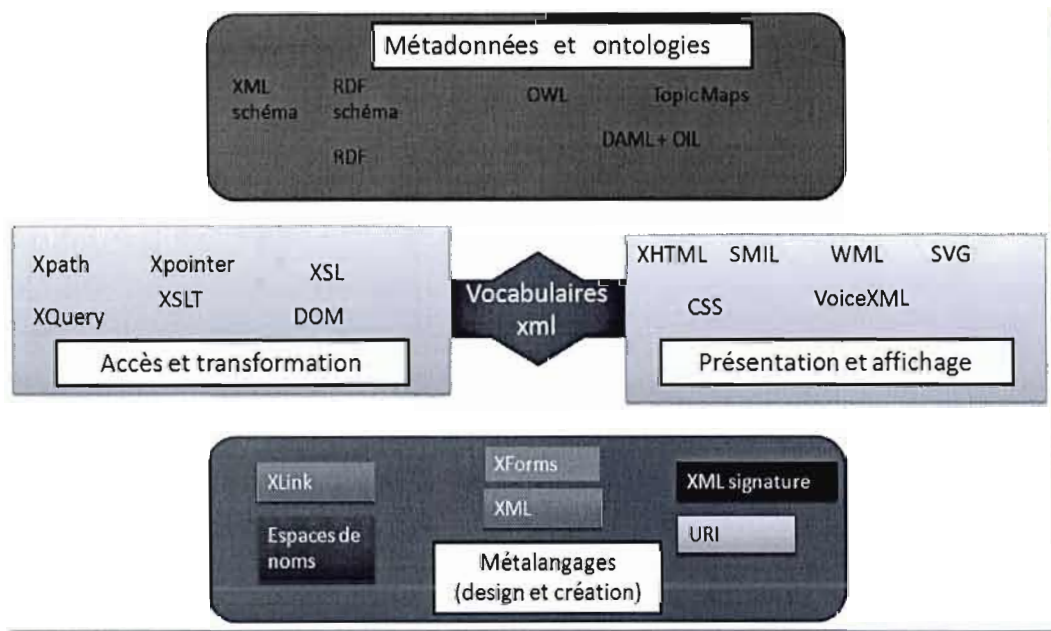


Figure 3.8 : Principales technologies de la famille XML (Geroimenko, 2006, p. 9)

XML est un langage de balisage permettant de structurer des données ou des documents sur le Web. Il est donc interopérable du point de vue syntaxique (Ta Tuan, 2005). À ce titre, il est considéré comme un standard de transport de données sur le Web. XML est en réalité un métalangage; il permet aux utilisateurs de définir leurs propres balises, d'ajouter une structure arbitraire à leurs documents sans rien dire de la signification sémantique de ces structures (Ta Tuan, 2005).

Les espaces de noms, identifiés chacun par un URI, permettent d'éviter les problèmes de collision de noms. Ce mécanisme permet d'apporter des extensions à un schéma en déclarant un nouveau vocabulaire et de mélanger ainsi dans un même document des éléments ou attributs dont les noms sont définis par différents vocabulaires. La description d'une ressource peut alors s'appuyer sur plusieurs schémas de métadonnées.

```
<Employe>Jean Dupont</Employe>  
<Professeur>Guy Tremblay</Professeur>  
<Cours>  
  <sigle>INF6560</sigle>  
  <EnseignePar>Guy Temblay</EnseignePar>  
</Cours>
```

En dépit de ses nombreux avantages, XML a ses limites eu égard aux besoins du Web sémantique. Pour les machines comme pour les humains, les balises XML ne sont

pas dotées de signification explicite. La balise **<Employe>**, si elle peut être interprétée par un utilisateur humain <sup>(36)</sup>, représente, aux yeux d'une machine, une simple chaîne de caractères. Ainsi, dans la portion du document XML ci-dessus, la requête « Donnez-moi la liste des employés » repérera uniquement Dupont <sup>(37)</sup>. Il s'agit évidemment d'une réponse incomplète puisque *Guy Tremblay* ne sera pas repéré alors que les professeurs sont également des employés ! Il faut donc trouver d'autres moyens permettant de décrire le contenu sémantique d'une telle balise. Deux solutions sont en concurrence : d'un côté le couple RDF/OWL supporté par le W3C et de l'autre, les *topic maps*, un standard proposé par l'ISO.

### 3.3.3. Couche RDF

Un des buts du Web sémantique est la recherche efficace d'information qui repose sur des métadonnées. Il est donc normal que « l'un des maillons du Web sémantique soit le langage RDF destiné à décrire les ressources » (Euzenat, 2003). L'objectif de la couche RDF est de permettre l'annotation structurée de ressources. Au lieu de définir un vocabulaire adapté à un usage précis comme l'avait fait le Dublin Core Metadata Initiative, le W3C a défini un cadre général pour la description des ressources du Web. Le RDF, *Resource Description Framework*, devenu une recommandation W3C en février 2004 est un modèle générique qui n'impose ni syntaxe ni vocabulaire.

---

<sup>36</sup> Sémantique implicite

<sup>37</sup> Cette question correspond à la requête XPATH //Employe



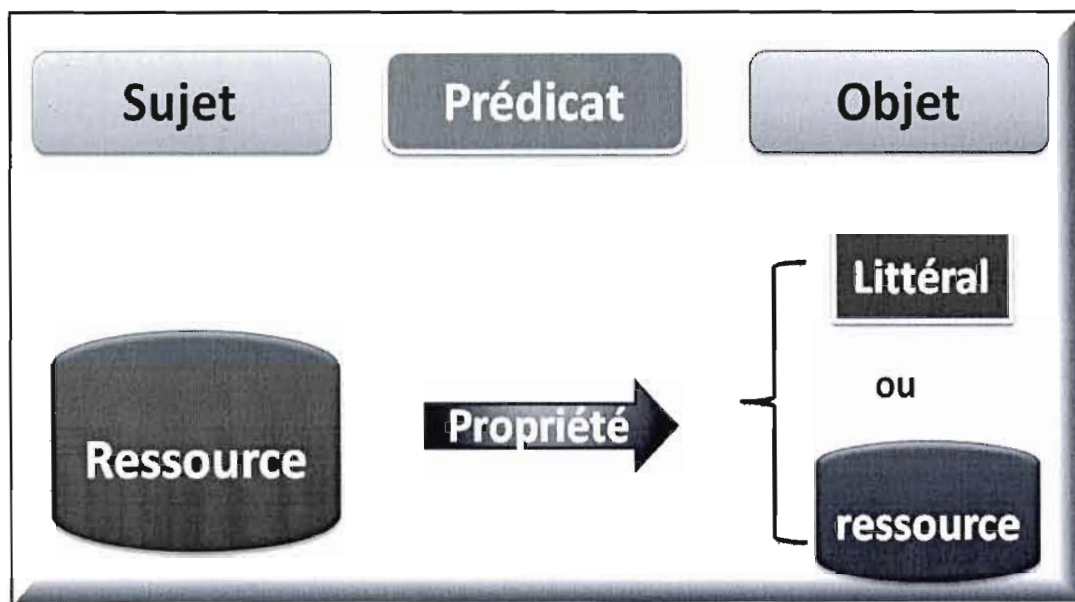


Figure 3.9: Le modèle RDF (Lacy, 2005, p. 81)

Ce modèle définit trois types d'objets : les ressources, les propriétés et les valeurs (figure 3.9). Les ressources peuvent être aussi bien des pages Web que tout objet ou personne du monde réel. Elles sont identifiées par leur URI, Uniform Resource Identifier. Les propriétés sont des attributs, des aspects, des caractéristiques qui s'appliquent à une ressource et peuvent prendre des valeurs particulières. La valeur d'une propriété peut être un littéral ou une autre ressource. Chaque propriété possède une signification précise, définit ses valeurs permises, les types de ressources qu'elle peut décrire et les relations qu'elle a avec les autres propriétés. Ces trois types d'objets peuvent être mis en relation par des déclarations ou assertions RDF, c'est-à-dire des triplets liant une ressource, une propriété et la valeur de cette propriété. Ces assertions sont elles-mêmes considérées comme des ressources. Dans une déclaration, la propriété est appelée *prédicat* tandis que la ressource décrite est le *sujet* de la description; la valeur de la propriété est appelée *objet*.



RDF offre quelques avantages. Premièrement l'identification universelle des ressources par un URI permet d'unifier facilement des descriptions de ressources provenant de différents sites distribués. En fonction de leurs besoins, les utilisateurs vont « pêcher dans divers vocabulaires dominants en s'appuyant sur une conceptualisation explicite d'un domaine de la réalité. Et, à partir du moment où l'élément de base des métadonnées est le triplet RDF, la rivalité entre les vocabulaires de métadonnées s'estompe car on peut les mixer ou expliciter les relations qu'ils entretiennent » (Roussel et al., 2007) tout en s'assurant que les concepts utilisés renvoient à une définition bien précise.

Deuxièmement, dans la mesure où une de ses syntaxes est définie en XML, RDF est une application XML. L'utilisation d'une syntaxe XML présente l'avantage de permettre l'échange de données. RDF et XML sont donc complémentaires ; le premier repose sur le second pour résoudre des problèmes d'interopérabilité syntaxique tels que l'encodage des caractères, l'internationalisation, etc.

Troisièmement, RDF n'est pas un schéma de métadonnées car il ne fournit pas une liste d'éléments à utiliser; c'est plutôt un modèle permettant de décrire les relations entre différentes ressources (Nogueras-Iso et al., 2004, p. 617). Tout comme XML, le modèle RDF n'impose pas le nom des balises qui seront utilisés. Au contraire, n'importe qui peut définir son propre vocabulaire. Quels sont les attributs autorisés? À quelles ressources s'appliquent-ils ? Quelles sont leurs valeurs admises? C'est le schéma RDF <sup>(38)</sup>, une recommandation du W3C depuis 2004, qui précise tous ces points. Le schéma RDF (ou RDFS) définit le vocabulaire utilisé dans une description RDF. On peut donc imaginer à loisir de nombreux schémas différents, adapté chacun

---

<sup>38</sup> RDF Vocabulary Description Language 1.0: RDF Schema (<http://www.w3.org/TR/rdf-schema/>).

à un domaine, une application spécifique ou un secteur d'activités en particulier tel que celui de l'enseignement supérieur.

Le langage RDFS permet d'enrichir RDF de quelques types de ressources prédéfinies (Euzenat, 2003); il définit des constructeurs de base permettant de définir des vocabulaires spécialisés. Son principe est simple : « il existe des **classes** de **ressources** et des **propriétés**. On peut définir des hiérarchies de classes et de propriétés dont l'applicabilité (*domain*) et les valeurs permises (*range*) peuvent être contraintes » (Ta Tuan, 2005). La propriété « *range* » permet de spécifier des restrictions sur les valeurs qu'une propriété peut prendre tandis que la propriété « *domain* » permet de préciser à quel type de ressources une propriété donnée peut être associée.

RDFS s'inspire de la programmation orientée objet. Cependant, contrairement au modèle objet, dans le cas du RDF, les propriétés et les classes sont définies séparément. On peut ainsi déterminer les classes pour lesquelles certaines propriétés seront utilisables. L'avantage de cette approche est que n'importe qui peut créer une propriété et désigner la classe à laquelle elle sera reliée (Hori, 2003).

Dans l'exemple de la figure 3.10, **Book**, **Writer**, **FamousWriter** sont définies comme des classes. La propriété « **hasWritten** » a été définie pour s'appliquer aux ressources de type « **Writer** » et ses valeurs admissibles sont de type « **Book** ». Dans la deuxième partie de la figure, une instance particulière est définie à l'aide de ce vocabulaire.

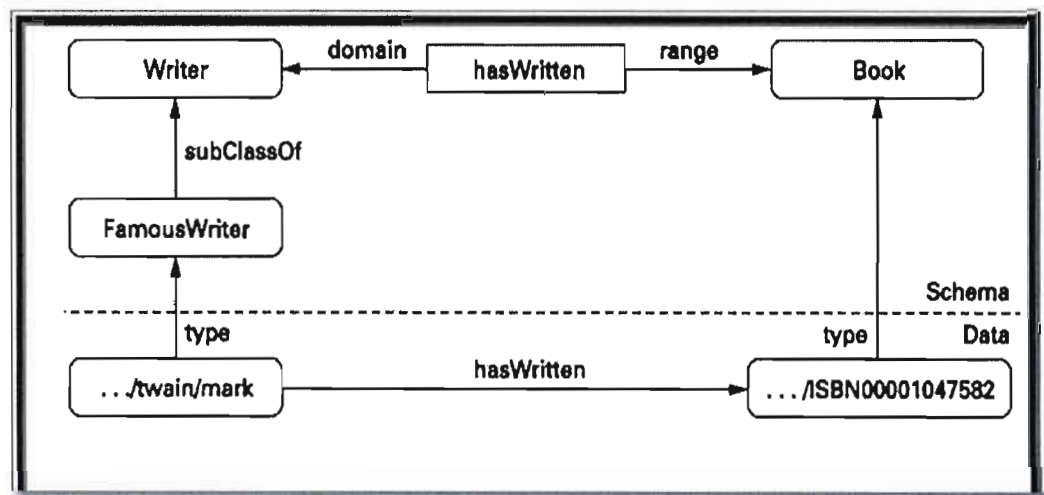


Figure 3.10 : Exemple d'un schéma RDF (Hori, 2003)

RDFS est donc un outil complémentaire à RDF. Alors que ce dernier est simplement un modèle de description, RDFS tente aussi de décrire la sémantique. RDFS est plus puissant que RDF car il fournit un moyen de construire un modèle objet qui nous permet de décrire le sens des objets (Hori, 2003). Dans la mesure où RDFS définit une taxonomie, c'est-à-dire une hiérarchie des classes, il facilite l'inférence et la recherche. Grâce à ses capacités de décrire les relations entre les ressources, RDFS ajoute une couche sémantique aux données contrairement à XML qui se limite à la syntaxe. Bien que RDFS n'intègre pas en tant que telles, des capacités de raisonnement, il fournit assez de sémantique pour que les applications puissent leur donner une interprétation opérationnelle (Laublet et al., 2003, p. 64). Néanmoins, la puissance d'expression de la signification des données n'est pas suffisamment puissante pour répondre entièrement aux besoins du Web sémantique. Par exemple, Hori (2003) reproche à RDFS l'incapacité de définir les propriétés relatives à une propriété, (2) les conditions nécessaires et suffisantes pour appartenir à une classe donnée, (3) les relations d'équivalence, de disjonction, etc. En plus, les seules contraintes exprimées par RDFS sont le domaine et les valeurs admissibles. Une

couche supplémentaire paraît donc nécessaire. Elle est basée sur l'usage d'ontologies lesquelles permettent de définir la sémantique des éléments de métadonnées (Nogueras-Iso et al., 2004).

### **3.3.4. Les ontologies**

Dans le Web sémantique, les ontologies jouent un rôle extrêmement important. N. Spivack (2006) les considère d'ailleurs comme « une forme de middleware sémantique » du Web. Elles font partie de nombreux formalismes de représentation de la connaissance tels que les vocabulaires contrôlés, les taxonomies, les thesaurus. D'ailleurs, la différence entre eux est souvent floue.

Un vocabulaire contrôlé est « une liste de termes, de mots ou d'expressions normalisés, utilisés pour indexer ou analyser le contenu et pour rechercher de l'information dans un domaine d'information défini. Il se caractérise par un format et une syntaxe uniformes, et peut contenir des synonymes et des renvois. Dans un vocabulaire contrôlé, un seul des termes possibles peut être utilisé pour représenter un concept. Par conséquent, toutes les ressources d'un corpus qui portent sur ce concept ou qui s'en approchent sont indexés à l'aide de ce terme représentatif » (Canada, 2005, p. 4). Un vocabulaire contrôle peut prendre plusieurs formes : d'une simple liste de mots ou de termes à un thésaurus « constitué d'un grand nombre de termes préférés et de termes rejetés, y compris des relations hiérarchiques et autres relations sémantiques entre les termes » (Canada, 2005, p. 5). Le thésaurus donne, en plus, de l'information sur les sujets connexes par l'ajout des termes associés. Dans le cas d'une taxonomie, le vocabulaire contrôlé est organisé sous forme hiérarchique simple. Les taxonomies et les thesaurus définissent des relations simples entre les termes mais ces relations ne peuvent pas atteindre la richesse

expressive des ontologies (Taniar & Rahayu, 2006). Ces différents formalismes offrent différents niveau d'expressivité sémantique comme le montre la figure 3.11 (Cardoso, 2007).

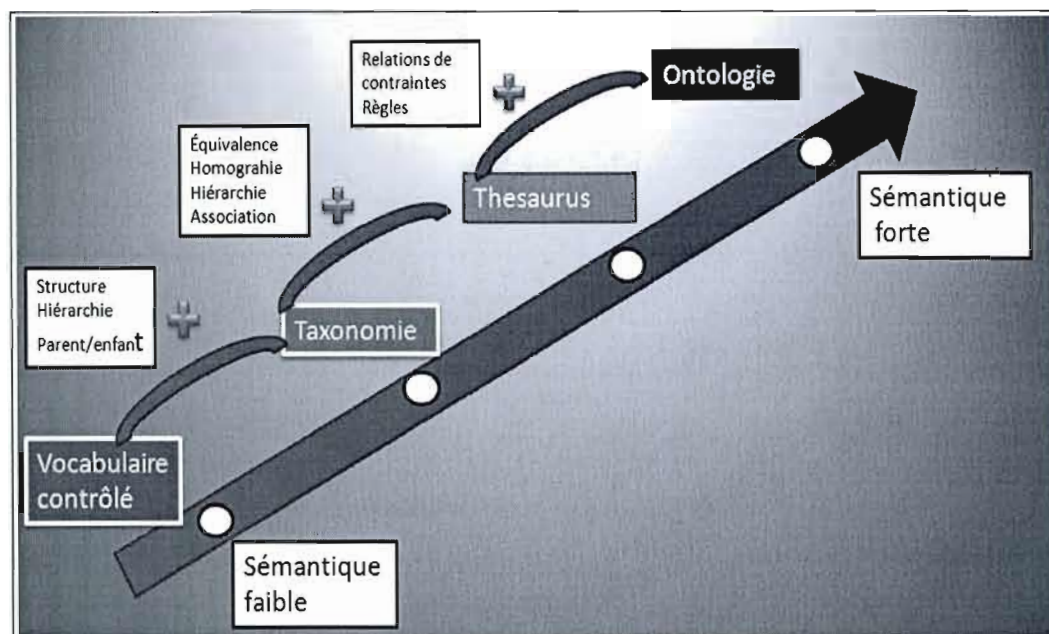


Figure 3.11 : Vocabulaires et richesse sémantique (Cardoso, 2007)

Utilisé depuis longtemps en philosophie, le terme « ontologie » a été récupéré par le domaine de l'informatique au début des années 90. Gruber (1993) est le premier à avoir proposé une définition qui fait désormais autorité : « une ontologie est une *spécification formelle et explicite d'une conceptualisation partagée* ». Cette courte définition, en apparence très simple, est riche en contenu. Nous ferons appel à Ghafour (2003) pour expliciter cette définition. *Conceptualisation* réfère à un modèle abstrait du monde réel et qui identifie les concepts de ce monde réel. *Formel* implique que la machine doit être capable d'interpréter la sémantique de l'information fournie. *Explicite* indique que les concepts utilisés et les contraintes de leur utilisation sont explicitement définis. *Partagée* parce qu'une ontologie exprime

une connaissance consensuelle, acceptée par un groupe, une communauté d'intérêt. Cela sous-entend qu'il serait illusoire de vouloir modéliser l'univers au complet mais qu'il faut se limiter plutôt à un domaine de connaissance en particulier. Ainsi, une ontologie, comprise comme un système structuré de concepts fondamentaux d'un domaine de connaissance, « peut jouer le rôle de médiateur grâce auquel les gens peuvent découvrir des différences ainsi que des points communs en ce qui concerne leur compréhension du monde cible » (Mizoguchi & Bourdeau, 2004).

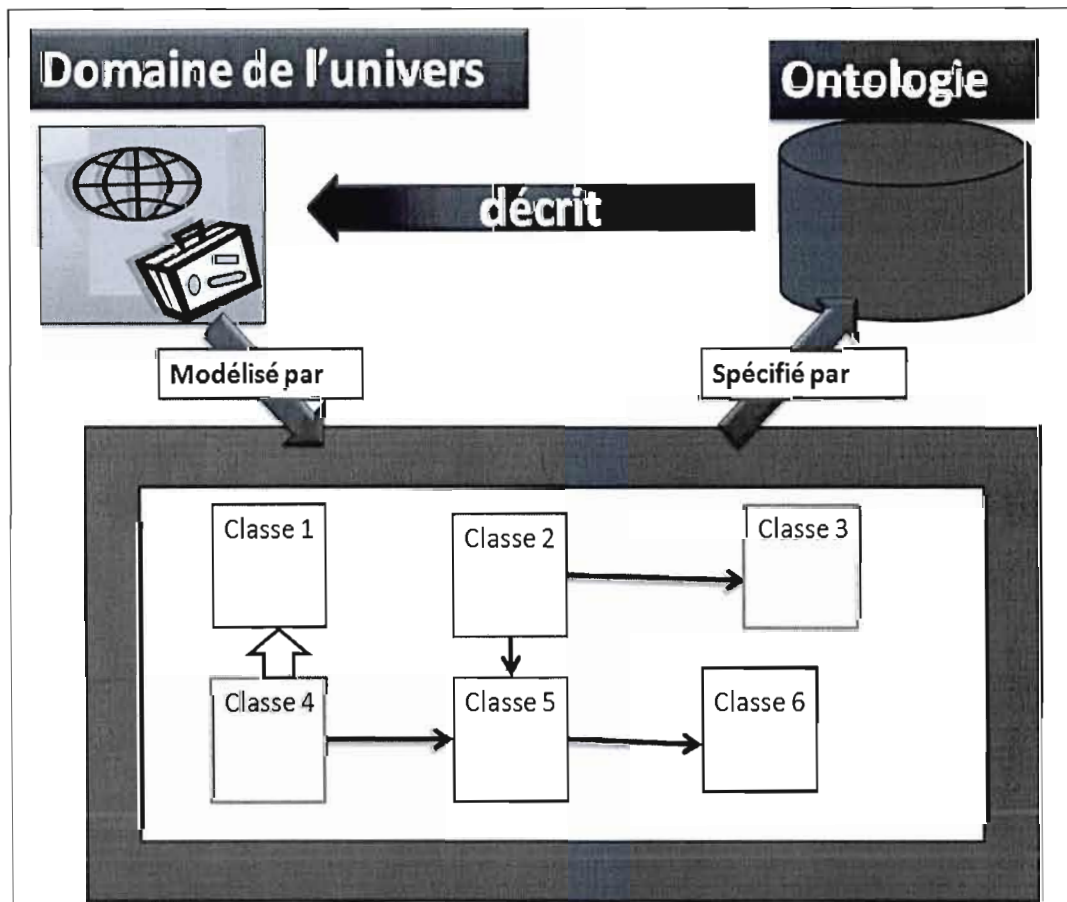


Figure 3.12 : L'ontologie selon Gruber (Lacy, 2005)

Une ontologie doit posséder les caractéristiques fondamentales suivantes : elle doit définir les concepts du domaine, leurs propriétés ainsi que leurs relations (figure 3.12). Ensuite, elle doit être fondée sur des principes formels rigoureux à partir desquels il est possible d'effectuer des raisonnements complexes (Golbreich et al., 2003). En fait, un thésaurus ou même une taxonomie sont des formes d'ontologie dont la grammaire n'a pas été formalisée et auxquelles on a ajouté des règles d'inférence tel que la transitivité ou la symétrie (Alesso & Smith, 2005). L'inférence est un processus qui utilise des règles pour manipuler la connaissance en vue de produire de nouvelles connaissances et choisir une série d'actions (Alesso & Smith, 2005, p. 34).

L'objectif premier d'une ontologie est de décrire de façon explicite la sémantique des données. Le but n'est pas de décrire le sens des termes de façon que l'ordinateur puisse le comprendre comme l'être humain ; il s'agit plutôt de définir les catégories simples des concepts et de relier chaque concept aux autres. La sémantique est une description formelle des termes et des relations entre eux afin de permettre aux machines d'interpréter l'information, de réduire l'ambiguïté de l'information et surtout d'assurer l'interopérabilité. Elle est fondée sur les ontologies spécifiées explicitement dans un langage de représentation suffisamment expressif pour représenter le sens des données tout en supportant un mécanisme d'inférence. Une ontologie est donc un vocabulaire contrôlé et organisé avec une formalisation explicite des relations entre les différents éléments du vocabulaire.

Cette formalisation exclut donc le langage naturel mais requiert plutôt un langage particulier. Selon Ta Tuan (2005), RDFS dispose de constructeurs permettant de représenter des ontologies simples sans capacités de raisonnement intégrées dans ce modèle. Cette idée est partagée par Hori (2003) pour qui RDFS s'approche des



langages de définition d'ontologies mais que sa sémantique reste limitée. L'ontologie a besoin des constructeurs de haut niveau que RDFS n'a pas. Il est nécessaire de spécifier par exemple, que des classes sont disjointes ou qu'une propriété est transitive. De plus, une ontologie peut contenir des contraintes reliant des concepts. Aussi, pour représenter les ontologies, le W3C a développé le langage OWL. Construit sur RDFS, il est plus expressif que ce dernier.

Le langage OWL est capable de répondre aux besoins de représentations d'une ontologie et de représenter explicitement le sens des termes et les relations entre ces termes. OWL a des moyens plus puissants pour exprimer la sémantique que XML, RDF et RDFS (Ghafour, 2003) : ce sont les classes, les propriétés et les instances. Le concept de classe est semblable à celui de la programmation objet et représente les objets du monde réel. Une classe représente un ensemble d'objets ayant des propriétés semblables. Le concept d'instance a la même signification que dans la programmation orientée objet sauf que dans le cas des ontologies, les instances sont exclusivement des représentations de l'information et n'ont pas de fonctionnalités ou méthodes associées. Les instances peuvent représenter des objets physiques ou virtuels. Le concept de propriété associe une paire attribut/valeur aux instances de classes. C'est donc une association binaire qui relie un objet à une valeur. La valeur peut être une donnée simple ou un autre objet. Une propriété peut être associée à différentes classes qui n'ont aucun rapport entre elles. Ce qui permet la réutilisabilité des propriétés. L'ontologie permet également d'explicitement les relations entre différents objets : les relations de synonymie (signification identique) et d'équivalence (valeurs identiques), les relations d'antonymie, les relations hiérarchiques (spécialisation et généralisation), les relations d'agrégation.



Grâce à ces caractéristiques, les ontologies peuvent être utilisées comme une forme de middleware sémantique. En effet, pour rendre l'échange des données efficace, il est essentiel de prendre en compte leur sémantique (Ghafour, 2003). La plupart des schémas de métadonnées sont définies avec une sémantique propre à leurs concepts. Ces schémas sont en quelque sorte des mondes fermés dont les concepts sont, à priori, incompatibles avec les concepts d'autres schémas. Pour assurer une interopérabilité entre différents schémas, il est nécessaire de pouvoir indiquer qu'un concept est équivalent à un autre concept appartenant à un schéma différent. Le couple RDF/OWL est idéal pour représenter les correspondances sémantiques entre des schémas différents. L'utilisation des ontologies permet de régler des problèmes de synonymie, d'équivalences sémantiques, de conflits de noms ou homonymie.

### **3.4. Le Web sémantique : fiction ou réalité ?**

Restreindre le Web sémantique à ses aspects techniques ou architecturaux serait trop limitatif (Laublet et al., 2003). Au-delà des différentes technologies qui le soutiennent et auxquelles les sections précédentes ont été consacrées, ce sont des applications développées qui feront de cette vision du Web une réalité. Si l'on en croit les défenseurs du Web sémantique, on devrait s'attendre à ce que l'adoption du Web sémantique se fasse sans aucune difficulté. Certains, plus sceptiques, soutiennent que pour des raisons psychologiques, politiques, commerciales et sociales, le Web sémantique sera confiné à un projet de recherche universitaire. Ils font valoir que depuis que Tim Berners-Lee a lancé cette idée, peu de progrès ont été réalisés, peu d'applications ont été développées et que finalement le Web sémantique reste marginal et limité principalement aux applications universitaires ! Plusieurs raisons sont à la base de cette situation. Les principales limites viennent de l'immaturité des outils technologiques actuellement disponibles et qui de surcroît,

sont d'une utilisation complexe. Ces outils sont conçus pour les experts en informatique spécialisés en représentation des connaissances ou en intelligence artificielle. L'autre problème est celui des ontologies. Actuellement, il y a peu d'ontologies car il est difficile de les concevoir. Les experts d'un domaine n'ont pas nécessairement les compétences en représentation des connaissances et ne connaissent pas le langage OWL. Ce manque de compétence cantonne donc les applications du Web sémantique entre les mains d'experts.

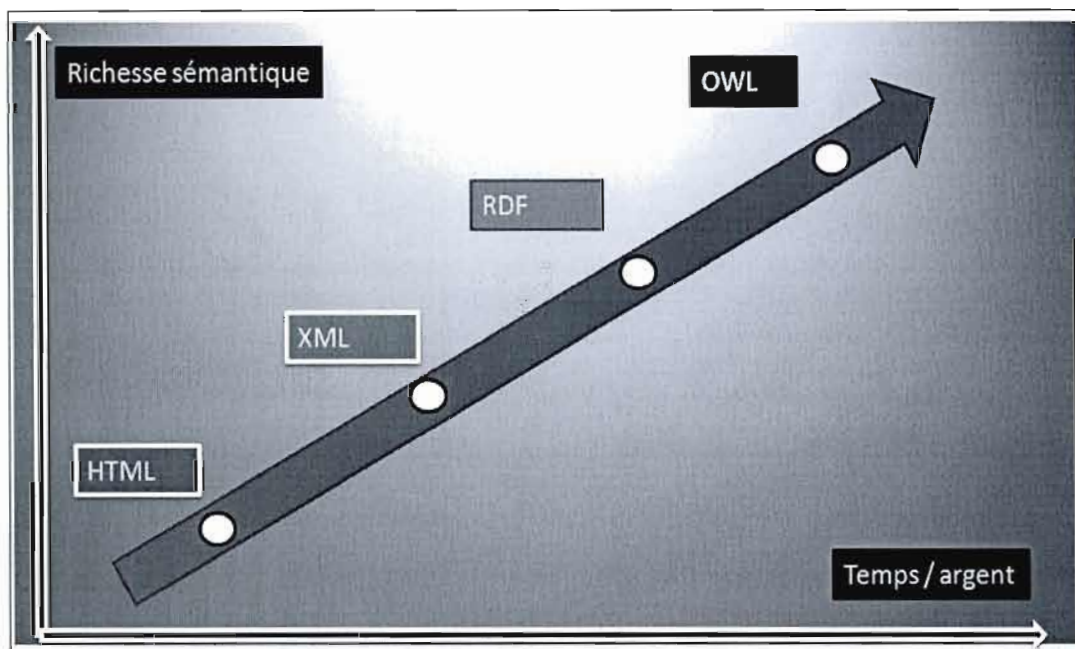


Figure 3.13 : Outils d'annotation et coûts associés, adaptée de Obrst (2003)

Un autre obstacle a trait aux coûts (figure 3.13). « Si la problématique initiale du Web sémantique est d'offrir à des agents logiciels la capacité d'exploiter le Web de manière intelligente, cette capacité n'est possible qu'à condition que les domaines de connaissances aient été décrits par des acteurs humains, qu'ils soient des spécialistes du domaine lors de la constitution d'ontologies de référence ou des auteurs, voire lecteurs actifs des documents, lors de leurs activités d'annotation à

l'aide des métadonnées faisant appel ou non à des ontologies préexistantes » (Caussanel et al., 2002). La mise en place du Web sémantique implique donc un travail important d'annotation du Web en ajoutant aux ressources des balises sémantiques.

Le balisage sémantique par des êtres humains et le développement d'ontologies posent un sérieux problème de coûts (Bouquet et al., 2006). Certains soutiennent même qu'il faudrait plutôt investir dans le développement des moteurs de recherche pour qu'ils soient plus aptes à mieux exploiter les pages Web tels qu'ils existent maintenant (Yonnet, 2007) ! D'autres soutiennent qu'une exploitation abusive du balisage sémantique à des fins de publicité ou de référencement de sites Web n'est pas exclue. Bref, pour toutes sortes de raisons, le Web sémantique apparaît pour certains comme une illusion, un concept purement théorique.

### **3.5. Vers le Web sémantique : les formats de transition**

Comme nous venons de le voir, un des problèmes des technologies du Web sémantique réside dans leur complexité. Deux solutions transitoires ont donc été envisagées afin d'utiliser XHTML pour ajouter des informations à valeur sémantique dans les ressources Web. La première solution fut basée sur les microformats, c'est-à-dire un ensemble de formats ouverts et standardisés destinés à ajouter des éléments sémantiques aux documents XHTML tandis que la deuxième repose sur les attributs RDF.

### 3.5.1. Les microformats

Les microformats sont une « tentative de combiner les principes du Web sémantique tels qu'annoncés par T. Berners-Lee avec les technologies actuelles du Web tel que XHTML » (Graf, 2007). L'idée des microformats est d'offrir un moyen simple permettant d'intégrer dans les ressources HTML des balises sémantiques. Nés du constat de complexité de la mise en œuvre des technologies RDF et OWL, les microformats sont « des attributs de classe particuliers, appartenant à une liste prédéfinie par une communauté. Ils sont ajoutés aux balises du code HTML, et jouent alors un double rôle de présentation de style et de structuration sémantique » (Dream'Orange, 2007, p. 1) des informations relatives aux événements, aux personnes, aux organisations, aux lieux et positions géographiques (géolocalisation), aux évaluations, aux CV, aux tags, etc.

Le projet « Microformats » <sup>(39)</sup>, initié en 2004 par Tantek Celik, a conçu plusieurs formats pour les besoins généraux comme l'information de contact, le calendrier d'événements, etc. On recense, à ce jour, plus de 80 microformats dont les plus connus sont hCalendar <sup>(40)</sup>, hCard <sup>(41)</sup>, hResume <sup>(42)</sup>, geo <sup>(43)</sup>, XFN <sup>(44)</sup>, hReview <sup>(45)</sup>,

---

<sup>39</sup> <http://microformats.org/wiki/>

<sup>40</sup> <http://microformats.org/wiki/hcalendar>

<sup>41</sup> <http://microformats.org/wiki/hcard>

<sup>42</sup> <http://microformats.org/wiki/hresume>

<sup>43</sup> Geo est un microformat destiné à définir un emplacement géographique, grâce à des coordonnées (<http://microformats.org/wiki/geo>)

<sup>44</sup> Le microformat XFN (XHTML Friends Network) a pour but décrire la relation entre les personnes par des liens hypertextes (<http://www.gmpg.org/xfn>).

rel-tag<sup>(46)</sup>, etc. Depuis peu, les éditeurs de contenus, les principaux acteurs du Web supportent le mouvement des microformats. Plusieurs extensions<sup>(47)</sup> permettent au fureteur Firefox de détecter les microformats et les prochaines versions des navigateurs dominants, entre autres Firefox 3 et IE 8, supporteront en natif les principaux microformats. De surcroît, en mars 2008, Yahoo a annoncé déjà qu'il allait indexer les microformats. Cette compatibilité des navigateurs avec les microformats offre la possibilité d'indexer les balises sémantiques ajoutées aux ressources du Web et de les rendre exploitables par diverses applications et services Web.

Ces microformats utilisent les éléments et attributs courants. Ils permettent d'inclure des aspects sémantiques dans les pages Web, tout en étant lisibles par les humains et par les machines. Cependant, même si les microformats sont des standards ouverts sous licence « Creative Commons », ils n'offrent pas l'extensibilité essentielle à l'interopérabilité (Adida, 2008). Par contre, ils ont largement inspiré l'approche basée sur les attributs RDF que nous allons détailler ci-dessous.

### **3.5.2. Les attributs RDF**

Les attributs RDF sont « un moyen d'incorporer des déclarations RDF dans une page Web et, à ce titre, sont similaires aux microformats » (Schapranow, 2006). Les deux

---

<sup>45</sup> Il s'agit d'un microformat destiné à décrire un avis donné par une personne sur à peu près n'importe quoi (un produit, un service, etc.) : (<http://microformats.org/wiki/hreview>).

<sup>46</sup> Fournit un moyen rapide d'associer des tags à une ressource (<http://microformats.org/wiki/rel-tag-fr>)

<sup>47</sup> L'extension Operator à Firefox de détecter la présence des microformats *adr*, *hCard*, *hCalendar*, *geo*, *tag* ou *rel – tag*, *xFolk* au sein du code XHTML.

méthodes ne sont pas concurrentielles. Au contraire, leur but final est le même : ajouter des métadonnées sémantiques dans une page Web. Elles ont d'ailleurs certaines caractéristiques en commun. Les métadonnées incorporées dans la page Web sont visibles et sont destinés aux utilisateurs humain et machine.

Mais il y a une différence importante : les microformats décrivent la ressource dans laquelle ils sont intégrés alors que les annotations RDF peuvent également décrire une ressource externe (figure 3.14).

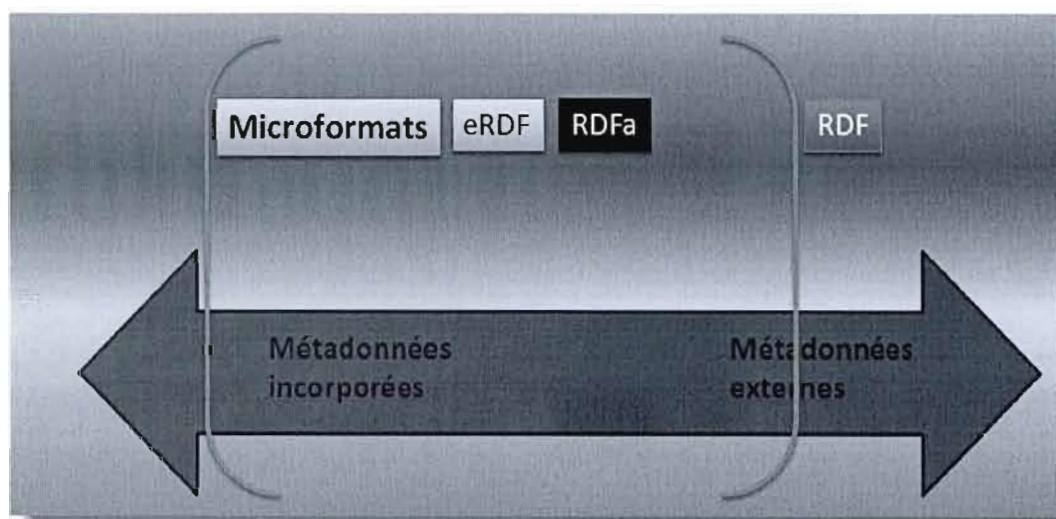


Figure 3.14 : Métadonnées internes et externes

Les deux approches permettent donc une adoption en douceur du Web sémantique (Schapranow, 2006). Mais force est de reconnaître que les microformats ne sont pas des recommandations du W3C et que leur niveau d'expressivité reste en dessous de celui des annotations RDF. Étant donné que les microformats correspondent à une liste d'attributs prédéfinie par une communauté, il faut une approche adaptable et flexible, reposant sur un jeu de règles générales pouvant s'appliquer à toute métadonnée (Allsopp, 2007). D'où l'intérêt des annotations RDF incorporées dans

une page XHTML et récupérées sous forme de triplets RDF. Dans cette approche, il existe deux variantes : eRDF (figure 3.15) et RDFa.



Figure 3.15 : Principes des balises RDF embarquées (Hori, 2003)

Embeddable RDF (eRDF) est le résultat d'une initiative personnelle <sup>(48)</sup>. Elle requiert l'indication du profil «<http://purl.org/NET/erdf/profile>» dans l'entête d'une page HTML. Le but de ce profil est de donner les instructions permettant de récupérer les triplets RDF. Pour insérer des expressions RDF dans une page XHTML, eRDF utilise les attributs HTML existants : 'rel', 'rev', 'class', 'href', 'src' et 'id'

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head profile="http://purl.org/NET/erdf/profile" >
<link rel="schema.dc" href="http://purl.org/dc/elements/1.1/" />
</head>
<body>
<h1 class="dc-title">Naked Metadata</h1>
<h2 class="dc-creator">Jonathan O'Donnell</h2>
<p class="dc-rights">http://purl.nla.gov.au/net/jod/tutorial/naked-metadata.html ©
Jonathan O'Donnell <span class="dc-date">23 October 2005</span></p>
</body>
</html>
```

Source : <http://jod.id.au/tutorial/naked-metadata.html>

<sup>48</sup> eRDF a été créé par Ian Davis en 2005 (<http://research.talis.com/2005/erdf/wiki/Main/RdfInHtml>)



Le schéma utilisé pour la description est indiqué par la balise « link ». Dans l'exemple ci-dessous, il s'agit du schéma Dublin Core.

**RDFa** <sup>(49)</sup>, proposé par le W3C (Adida et al., 2008), part du même principe : ajouter des informations dans le document HTML en réutilisant les propriétés déjà définies dans des vocabulaires RDF existants. Des mécanismes de transformation peuvent alors récupérer ces données pour l'écriture des triplets RDF de façon dynamique. Beaucoup plus riche que eRDF et piloté par le W3C, le RDFa <sup>(50)</sup> (Adida et al., 2008) utilise la syntaxe HTML existante (*content*, *src*, *rel*, *rev*, *href*) mais y ajoute de nouveaux attributs : *about*, *resource*, *typeOf*, *property*, *datatype*,. C'est pourquoi le W3C a mis au point une nouvelle DTD issue de XHTML et intégrant les attributs de RDFa :

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN"
"http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
```

Le RDFa utilise le principe des CURIs (URIs compactes) mis au point par le W3C pour simplifier l'écriture des URIs en exploitant les espaces de noms. À titre d'exemple, plutôt que d'écrire

```
<h1 property="http://purl.org/dc/elements/1.1/title">Web sémantique</h1>
```

on peut simplifier l'écriture ainsi :

```
<h1 property="dc:title">Web sémantique</h1>
```

---

<sup>49</sup> Anciennement RDF/A, RDFa est devenu une recommandation du W3C en octobre 2008 (<http://www.w3.org/TR/2008/REC-rdfa-syntax-20081014/>)

<sup>50</sup> RDF attributes



Pour cela, il faut avoir déclaré au préalable l'espace de nom correspondant au schéma utilisé. Le CURI a la forme suivante :

**Préfixe : clé**

L'usage de plus en plus répandu des microformats a amené le W3C à mettre au point une nouvelle recommandation, le GRDDL dont le but est de rendre les pages Web actuelles plus lisibles pour les applications du Web sémantique. Le GRDDL, *Gleaning Resource Descriptions from Dialects of Languages*, littéralement « glanage des descriptions de ressources à partir des dialectes de langages », est un mécanisme permettant d'extraire ou de glaner les informations présentes dans une page Web en format XHTML pour les transformer en triplets RDF.

Le GRDDL introduit une syntaxe permettant de déclarer qu'un document contient des données récupérables en RDF et de pointer vers un algorithme permettant cette récupération qui est, le plus souvent, un script XSLT. La figure 3.16 explique son mécanisme de fonctionnement (Halpin & Davis, 2007).

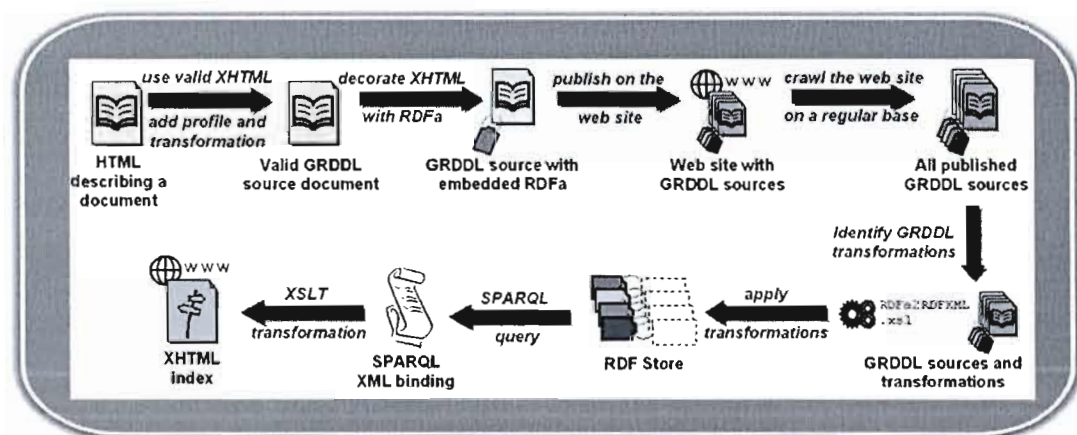


Figure 3.16 : Fonctionnement de GRDDL <sup>(51)</sup>

<sup>51</sup> <http://www.w3.org/TR/grddl-primer/>

Aujourd'hui, il y a d'ailleurs une convergence entre les microformats et les annotations RDF. Cette convergence est rendue possible grâce à GRDDL (figure 3.17).

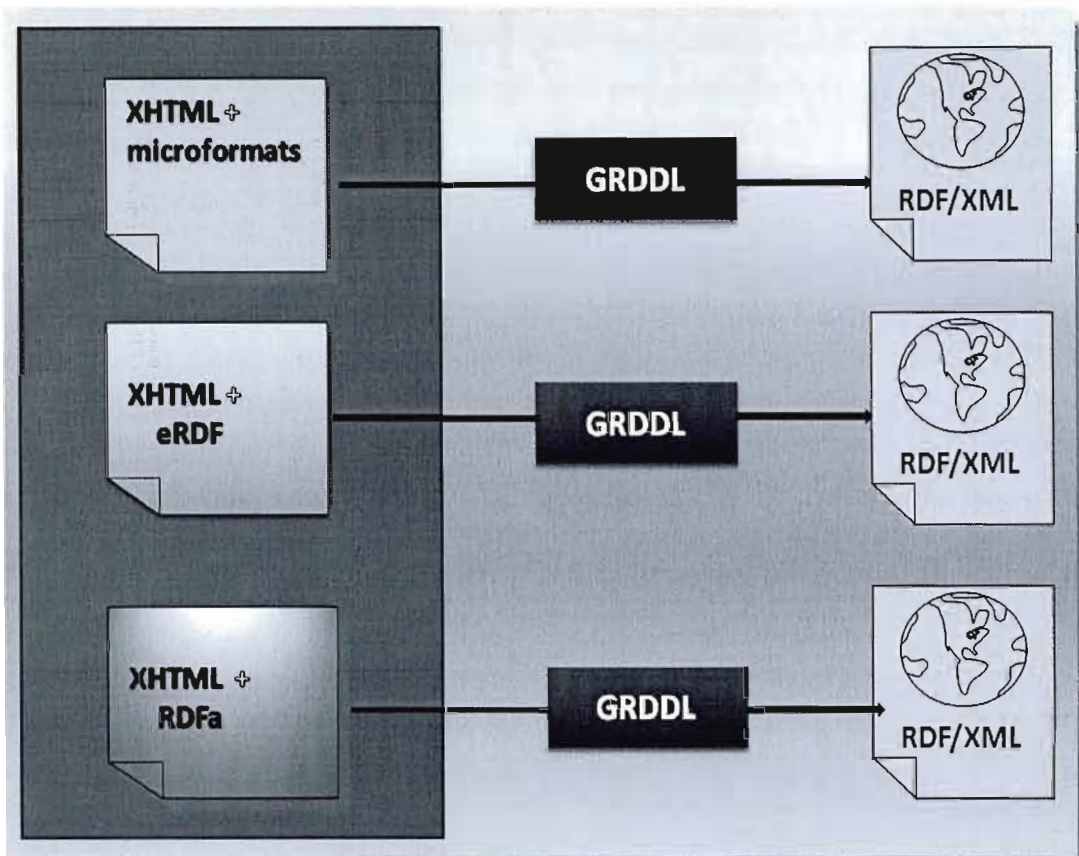


Figure 3.17: Convergence entre microformats et RDFa , adaptée de Adida (2008)

À partir d'un document XHTML, et grâce à GRDDL, on peut générer autant de documents RDF que de dialectes utilisés dans la page.

Même si les microformats ne font pas nécessairement l'unanimité, on reconnaît qu'ils favorisent la promotion du Web sémantique grâce à leur simplicité. En revanche, ils ne sont utiles que dans le cas de vocabulaire de description simple comme FOAF, iCal. Beaucoup moins puissants que RDF/OWL, ils ne pourront jamais se substituer à eux. C'est ce que David W. Cearley et ses collègues (Cearley et al., 2007) appellent l'hypertexte sémantique, c'est-à-dire un mécanisme permettant d'embarquer de la sémantique rudimentaire dans les documents HTML (figure 3.18).

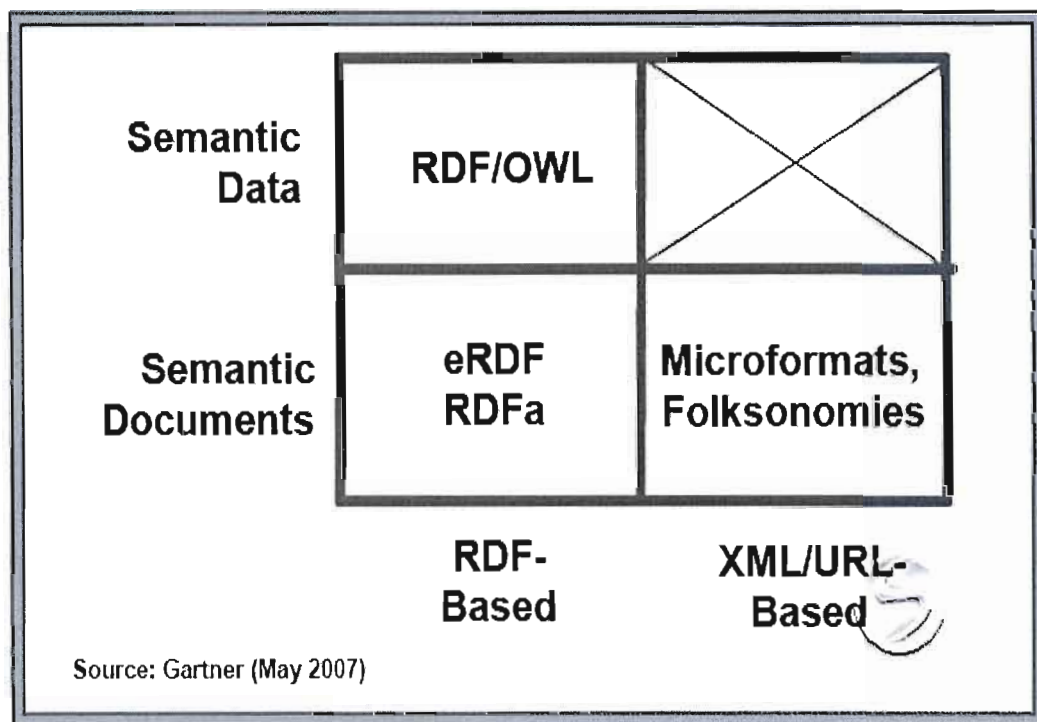


Figure 3.18: Hypertexte et Web sémantique (Cearley et al., 2007)

### 3.6. Le Web sémantique : perspectives

En dépit du bien fondé de certaines critiques adressées au Web sémantique, celui-ci ne tient plus de la science fiction si l'on en croit plusieurs études. Au cours des

dernières années, les choses ont bien évolué. Les technologies de base sont bien définies. Les entreprises s'y intéressent de plus en plus. La compagnie Adobe, par exemple, a adopté le format RDF pour ses différentes applications (Ball, 2007) et plusieurs sociétés ont mis au point des systèmes sophistiqués et très efficaces de balisage automatique des informations contenues dans une page Web et qui semblent donner des bons résultats (Yonnet, 2007).

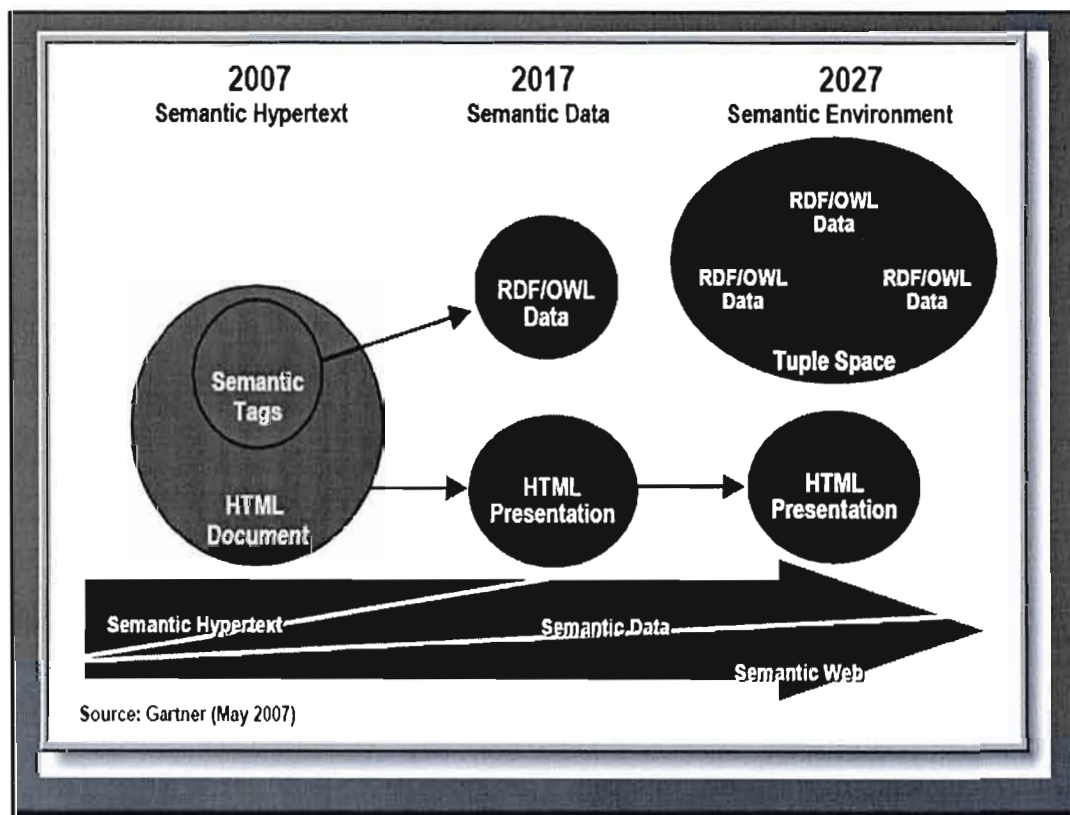


Figure 3.19 : Développement du Web sémantique (Cearley et al., 2007)

D'après une étude du groupe Gartner (Cearley et al., 2007), l'adoption des technologies et langages du Web sémantique est une réalité indéniable. Selon leurs prévisions (voir figure 3.19), en 2012, 70% des pages publiques seront dotées d'annotations sémantiques mais seulement 20% feront un usage intensif

d'ontologies, et vers 2017, le Web sémantique tel que défini par Tim Berners-Lee sera réalité. Ces prévisions sont confirmées par celle de N. Spinack (2007c) qui présente l'évolution du Web selon deux axes (la connectivité informationnelle et la connectivité sociale) à partir de l'ère préhistorique des ordinateurs personnels jusqu'à l'arrivée prochaine du Web sémantique. Pour lui, nous sommes à la fin du Web 2.0, considéré comme « le Web du peuple », et au début du Web 3.0 qui correspond au Web sémantique et qui devrait se généraliser pendant la décennie 2010 – 2020 (figure 3.20).

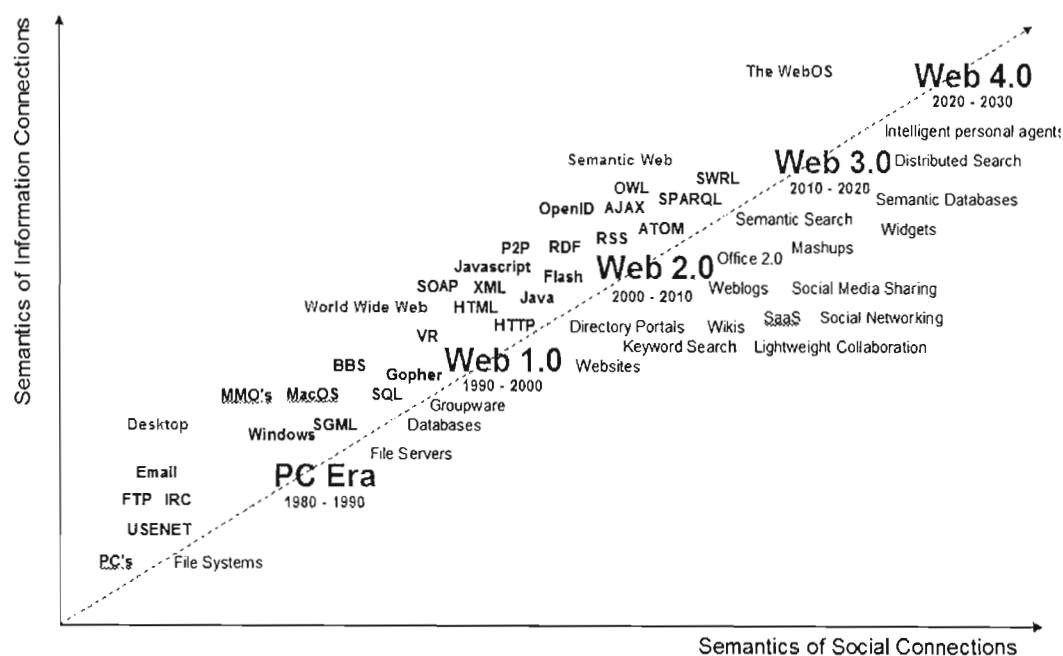


Figure 3.20 Évolution du Web selon N. Spinack (2007c)



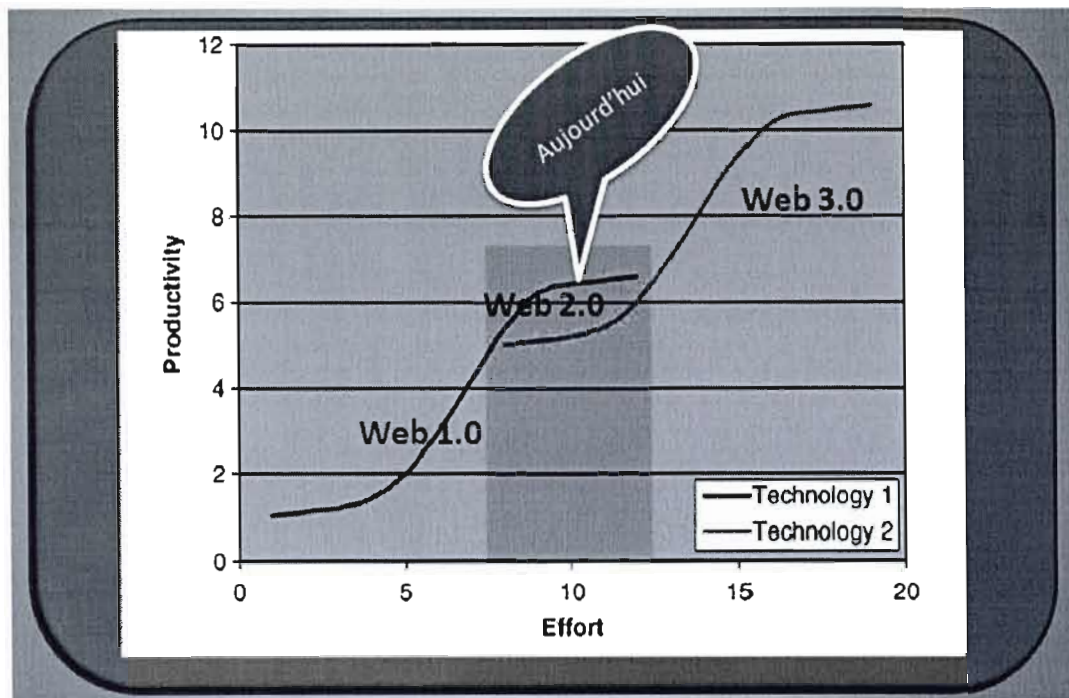


Figure 3.21 : Courbe en S de Richard N. Foster adaptée au Web sémantique

Comme toute technologie, il peut se passer des années entre le moment où elle a atteint un degré de maturité suffisant et le moment de son adoption par les acteurs du marché. Un des modèles de diffusion d'une nouvelle technologie est la courbe en S (figure 3.21) de Richard N. Foster (1986).

Selon cette courbe, tout indique que le Web sémantique a tout simplement besoin de ce qu'on appelle des « killer applications » pour prouver sa valeur et susciter son adoption rapide. La première génération du Web (Web 1.0) a eu Mozilla, Netscape, IE, HTML, YAHOO, GOOGLE ainsi que des milliers de sites Web populaires. Le Web sémantique (Web 3.0) a besoin de telles applications.



Figure 3.22 : Outils du Web sémantique en développement <sup>(52)</sup>

De nombreuses compagnies sont entrain de développer des outils tel que *Swoogle* (figure 3.22) pour la construction des applications et sites sémantiques et ces outils s'efforcent de cacher la complexité des technologies utilisées afin de permettre au commun des mortels de concevoir des applications et des contenus qui tirent profit de la puissance du Web sémantique sans avoir un PhD en représentation des connaissances (Spivack, 2006). Bref, selon Spinack, dans dix ans, le Web sémantique sera présent dans la plupart des sites et autres applications Web; dans 20, tout le contenu Web et applications sur Internet seront intégrés dans le Web sémantique (Spivack, 2007b).

---

<sup>52</sup> <http://swoogle.umbc.edu/>

## Deuxième partie

# Distribution des métadonnées dans le Web sémantique : le cas du *MLR* <sup>(53)</sup>

---

<sup>53</sup> ISO/IEC CD3 19788 : Information technology – Learning, education and training – Metadata for learning resources. Part. 1 : Framework; Part. 2: Core elements.



## Introduction

Une des fonctions du Web sémantique est la description, de manière formelle, du contenu des ressources qu'il contient afin que celles-ci soient facilement repérables. Dès lors, on comprend aisément que différents acteurs, et en particulier les producteurs et diffuseurs de ressources d'information, suivent avec beaucoup d'intérêt les développements du Web sémantique. La visibilité des ressources est fortement liée à leur indexation. En effet, cette visibilité tant recherchée n'est possible que si un travail préalable d'indexation ou d'annotation a été effectué. Dans ce contexte, développer des schémas de métadonnées qui répondent aux exigences du Web sémantique est un enjeu important.

Les chapitres précédents ont mis en lumière l'apport des technologies du Web sémantique à la solution du problème de l'interopérabilité des métadonnées. Nous avons vu que dans la réalité, l'utilisation de ces technologies reste limitée particulièrement à cause de leur complexité. Le but de cette deuxième partie est de montrer une approche pragmatique permettant de distribuer des éléments de métadonnées MLR et de les rendre disponibles aux applications du Web sémantique.

Le domaine utilisé comme cas d'étude est celui de l'éducation, pour une raison bien simple. Il s'agit d'un domaine où les travaux de normalisation ont atteint un niveau de maturité reconnu. Grâce à un travail soutenu de normalisation qui a atteint un point culminant en 2002 avec l'adoption du *Learning Object Metadata standard* (<sup>54</sup>), le domaine de l'éducation s'est doté de nombreux schémas de métadonnées.

---

<sup>54</sup> IEEE 1484.12.1-2002 : IEEE Standard for Learning Object Metadata (<http://ltsc.ieee.org/wg12/>)

L'existence des réseaux de communication efficaces, jumelée à la volonté de réaliser des économies d'échelle a incité le secteur de l'éducation à adopter une « approche plus globale pour, d'une part réduire la lourdeur d'adaptation et de mise à niveau du matériel pédagogique et, d'autre part, augmenter l'efficacité du repérage et du partage des ressources parmi les divers utilisateurs et contextes technologiques dédiés à la e-formation » (Oubahssi, 2005, p. 47). Un des objectifs visés par la normalisation est l'accessibilité des ressources en vue de leur utilisation et réutilisation. La question n'est donc pas de normaliser la formation ou le contenu, mais le mode de production et de description des ressources (SDTICE, 2007). Les progrès réalisés n'ont malheureusement pas permis d'aboutir à un schéma universellement accepté. La recommandation de «faire du LOM une norme internationale a été refusée par certains pays, en raison notamment, de la vision pédagogique sous-jacente » (SDTICE, 2007, p. 10). Pour répondre à des besoins socioculturels différents, plusieurs profils d'application du LOM ont été développés <sup>(55)</sup> sans que l'interopérabilité entre eux soit garantie. C'est, en particulier, pour répondre à cette préoccupation que le MLR a été pensé.

Nous avons donc choisi de consacrer les prochains chapitres à ce nouveau schéma de métadonnées afin de voir dans quelles mesures cette future norme ISO <sup>(56)</sup> va contribuer à l'amélioration de l'interopérabilité des métadonnées en s'intégrant plus rapidement que les normes qui l'ont précédé dans le Web sémantique. Cette partie sera divisée en deux chapitres. Au chapitre 4 nous présenterons le MLR et verrons particulièrement les raisons qui justifient une nouvelle norme de métadonnées dans

---

<sup>55</sup> Comme exemples de profils d'application du LOM, on peut citer, pour la France, le LOM-FR, et, pour le Québec, NORMETIC.

<sup>56</sup> Une fois adopté, le MLR aura comme numéro officiel ISO/IEC 19788

le domaine de l'éducation malgré le foisonnement de nombreux schémas au cours des dernières années. Le chapitre 5 présente une approche pragmatique permettant de distribuer les éléments de métadonnées MLR à travers le Web et la façon d'utiliser le MLR dans le cadre du Web 2.0. Dans ce chapitre, nous examinerons dans un premier temps, les atouts dont dispose le MLR pour favoriser une plus grande interopérabilité avec des schémas existants ou à venir et s'intégrer au Web sémantique.

## CHAPITRE IV

### METADATA FOR LEARNING RESOURCES <sup>(57)</sup>

Le partage des ressources d'enseignement et d'apprentissage est fortement lié à la capacité d'échanger des métadonnées qui les décrivent. En l'absence d'une norme internationale pouvant concurrencer le LOM (SDTIC, 2007, p. 10), celui-ci avait toutes les chances de s'imposer comme standard de facto dans le domaine de l'éducation. Mais, pour des raisons que nous détaillerons ci-dessous, plusieurs schémas de métadonnées sont actuellement utilisés pour indexer les ressources pédagogiques. Cette diversité des schémas est à la base du problème d'interopérabilité, comme nous l'avons souligné au chapitre 1.

#### 4.1. Pourquoi le MLR

Le LOM se définit comme un modèle standard de métadonnées permettant de décrire et référencer toute ressource d'enseignement et d'apprentissage. En ce sens, il fournit une base concrète de départ pour la normalisation et l'indexation des ressources d'enseignement mais son implantation ne fut pas aussi large qu'on l'aurait souhaité. Les résultats d'une vaste étude sur son implémentation <sup>(58)</sup> sont assez éloquents (ISO/CEI JTC1 SC36, 2004). Les éléments de données les plus utilisés

---

<sup>57</sup> ISO/IEC CD3 19788 : Information technology – Learning, education and training – Metadata for learning resources. Part. 1 : Framework; Part. 2: Core elements.

<sup>58</sup> L'étude portait sur trois points : les éléments de données retenus dans les profils d'application, les éléments de données effectivement renseignés et les valeurs utilisées pour ces éléments de données.

correspondent essentiellement aux éléments du Dublin Core si bien qu'un petit nombre d'éléments de données du LOM sont effectivement renseignés <sup>(59)</sup>. Tous les éléments du LOM sont optionnels y compris ceux ayant trait aux aspects pédagogiques alors qu'ils sont l'essence même du LOM. Par ailleurs, la complexité de la structuration du LOM (modèle hiérarchique) rend difficile la réutilisabilité de certains éléments (Friesen, 2004).

Malgré ses insuffisances et ses inconvénients, le LOM a été adapté dans le cadre de nombreux profils d'application. Dans l'esprit de la norme ISO/IEC 11179-3 <sup>(60)</sup>, la création d'un modèle conceptuel garantit l'interopérabilité des schémas de métadonnées issus de ce modèle. Ces derniers sont en quelque sorte considérés comme des instanciations d'un seul et unique modèle conceptuel; ce qui assure, en théorie, leur interopérabilité. Du point de vue de la norme ISO/IEC 11179-3, le LOM n'est pas un modèle conceptuel et cela, pour plusieurs raisons. D'une part, pour certains de ses éléments, le LOM définit la liste des mots en anglais qui sont utilisables comme valeurs permises ; ce qui est du ressort du niveau « représentation » de la norme. D'autre part, pour des éléments correspondants à des dates, le LOM définit la façon de les représenter. Enfin, le LOM est présenté sous forme de tableau, ce qui n'est pas le cas d'un modèle conceptuel (Delestre & Bourda, 2005). En matière d'interopérabilité, la conclusion de l'étude mentionnée ci-dessus (ISO/CEI JTC1 SC36, 2004) est que le LOM n'est pas un modèle à suivre.

---

<sup>59</sup> Alors que le Dublin Core spécifie seulement 15 éléments de données, le LOM en compte 76 pour décrire les différentes caractéristiques d'un objet pédagogique.

<sup>60</sup> ISO/IEC 11179-3 Information Technology -- Registry metamodel and basic attributes

D'autres critiques sont également formulées à l'égard du LOM par différents acteurs de la normalisation. Certains lui reprochent notamment « une certaine incohérence entre la définition générique des objets pédagogiques et les éléments permettant de les décrire, l'intégration au sein d'un même modèle, des entités de niveaux conceptuellement différents (les ressources nécessaires à la mise en place d'activités pédagogiques et les activités elles-mêmes), le fait que le problème de la signification des termes choisis et de la définition des métadonnées ne soit pas complètement résolu, etc. » (La Passardière & Jarraud, 2004) (Delestre & Bourda, 2005). D'autres lui reprochent le manque de précision dans la définition des éléments et des valeurs possibles (Guay, 2007). Une autre critique, non moins importante, est que le LOM a balayé sous le tapis bien des aspects multilingues et multiculturels. À titre d'exemple, dans le LOM, le nom d'un élément de données est confondu avec sa signification. Or celle-ci n'est pas neutre ; elle dépend du contexte, de la culture, de la langue.

Compte tenu de ces critiques et considérant particulièrement le manque d'interopérabilité des différents profils du LOM, l'idée d'élaborer une nouvelle norme pour la description du patrimoine éducatif capable d'améliorer le LOM fut mise de l'avant par le groupe de travail *GT4* du sous-comité SC36 de l'ISO/CEI. Nous allons consacrer la section qui suit à la présentation de cette norme, le *Metadata for Learning Resources*, dans son stade actuellement de développement, c'est-à-dire le CD3 <sup>(61)</sup> <sup>(62)</sup> (ISO/CEI JTC1 SC36, 2008a) (ISO/CEI JTC1 SC36, 2008b)

---

<sup>61</sup> Committee Draft, version 3

<sup>62</sup> Sur recommandation du WG4 lors de sa réunion de novembre 2008 à Paris, le comité éditorial a procédé à une révision majeure de ce document et a produit un "sneak peek FCD" le 19 mai dernier. Dans sa résolution no 11 lors de la réunion tenue à Wellington (Nouvelle-Zélande), le WG4 recommande aux rédacteurs de préparer pour la fin juin 2009, sur la base de ce « sneak peek », un projet de document FCD qui fera l'objet d'un examen lors de la réunion de septembre 2009 en Suède.

#### 4.2. Présentation du MLR

Le MLR est un projet de norme internationale pour la description des ressources d'enseignement et d'apprentissage. Son élaboration repose sur trois objectifs fondamentaux: l'adaptabilité culturelle et linguistique, la capacité d'extension permettant de répondre aux besoins particuliers de différentes communautés d'utilisateurs et le souci d'interopérabilité avec des schémas existants ou à venir (Guay, 2007).

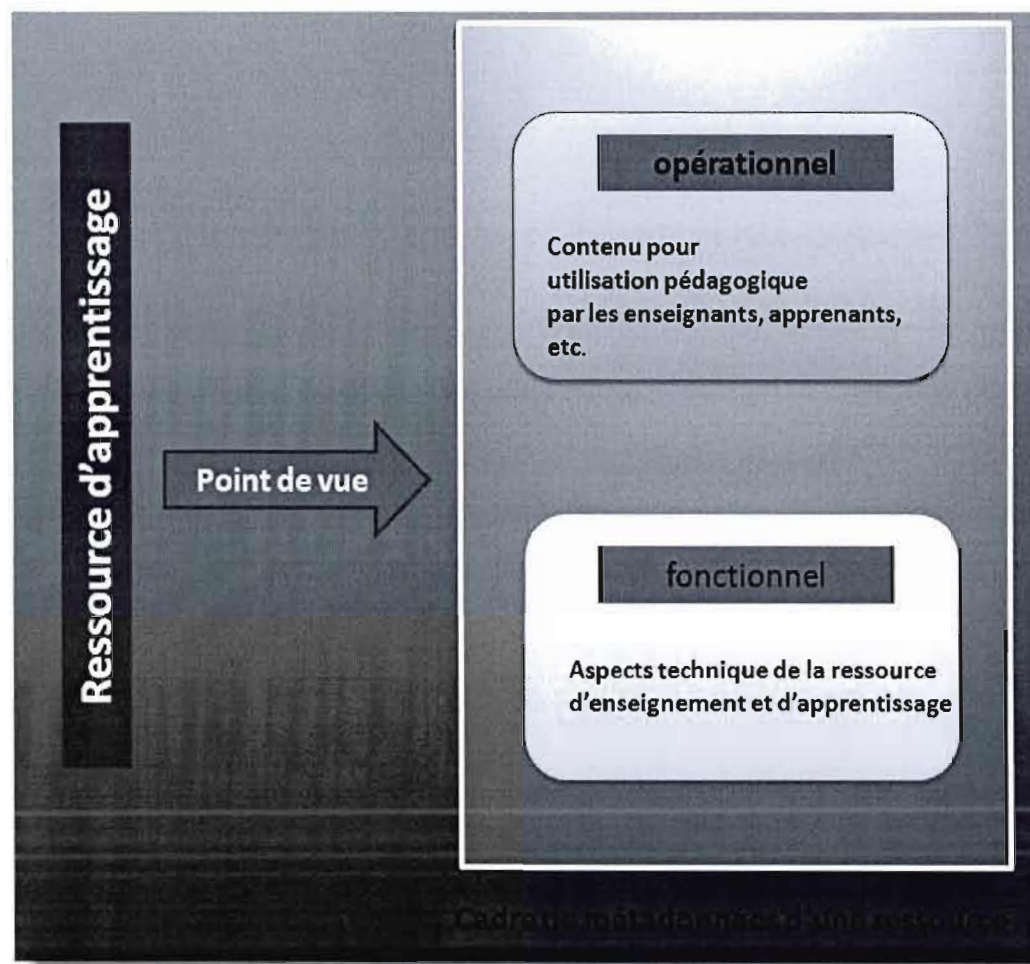


Figure 4.1 : Cadre de métadonnée d'une ressource (ISO/CEI JTC1 SC36, 2008a)

Le MLR fournit un cadre d'identification, de spécification, de représentation syntaxique et sémantique des métadonnées d'une ressource d'apprentissage en tenant compte de la diversité des contextes culturels et linguistiques dans lesquels les ressources et leurs métadonnées sont créées et exploitées. L'objectif du MLR est double : d'une part, fournir une approche standardisée permettant d'identifier et de spécifier les éléments de métadonnées nécessaires à la description d'une ressource d'enseignement et, d'autre part, faciliter la recherche, l'acquisition, l'évaluation et l'utilisation de ces ressources par les apprenants, les enseignants ainsi que par les systèmes logiciels. Elle vise ainsi à favoriser le partage et l'échange des ressources ainsi que des métadonnées qui leur sont associées. Dans l'esprit du MLR, une ressource d'apprentissage est toute entité qui peut être référencée par un URI et utilisée à des fins d'apprentissage, de formation ou d'éducation. Même si cette ressource peut être numérique ou non, la norme s'inscrit dans une perspective des ressources numériques et accessibles en ligne grâce au réseau Internet.

Le MLR est conçu dans une optique qui concilie deux points de vue différents. D'un côté, le point de vue opérationnel qui s'attache au contenu de la ressource et, d'un autre côté, le point de vue fonctionnel orienté vers les aspects techniques de la ressource (figure 4.1). Fondé sur le principe de la modularité, le MLR est une norme en plusieurs parties. Tandis que certaines parties seront guidées par le point de vue opérationnel, d'autres mettront l'accent sur les aspects fonctionnels selon les besoins à combler. Le nombre de parties n'est pas encore défini mais six sont déjà planifiées <sup>(63)</sup> :

ISO/IEC 19788-1 Metadata for Learning Resources – Part 1: Framework  
 ISO/IEC 19788-2 Metadata for Learning Resources – Part 2: Core Elements

---

<sup>63</sup> La norme prévoit également des parties consacrées aux différents « bindings » tel que XML, RDF, etc.



ISO/IEC 19788-3 Metadata for Learning Resources – Part 3: MLR Core Application Profile  
ISO/IEC 19788-4 Metadata for Learning Resources – Part 4: Technical Elements  
ISO/IEC 19788-5 Metadata for Learning Resources – Part 5: Educational Elements  
ISO/IEC 19788-6 Metadata for Learning Resources – Part 6: Availability and Rights Management

Au stade actuel des travaux, deux parties font l'objet du processus d'acceptation comme norme ISO : la partie 1 qui constitue le cadre de la norme et qui établit la structure d'une notice de métadonnées propres à une ressource pédagogique et la partie 2 qui définit un ensemble restreint d'éléments de base.

#### **4.2.1. Le Framework**

Cette partie de la norme précise les principes fondamentaux sur lesquels elle s'appuie. Elle précise également les structures de données utilisées et les règles suivies pour la spécification des métadonnées nécessaires à la description des ressources concernées. De plus, elle spécifie les règles à suivre pour l'attribution et la gestion des identifiants attribués à chacune des spécifications des éléments de métadonnées ainsi que pour le développement des autres parties de la norme. Les règles pour la spécification, la construction et le développement des vocabulaires contrôlés sont également détaillées. Le *Framework* fournit également des cartes conceptuelles <sup>(65)</sup> qui représentent graphiquement les concepts du MLR. La figure ci-dessous représente la carte conceptuelle d'une spécification d'un élément de métadonnées.

---

<sup>65</sup> Une représentation graphique dans laquelle les concepts sont liés entre eux par des liens pour former un réseau sémantique et qui facilite la compréhension et l'appropriation d'une réalité complexe.

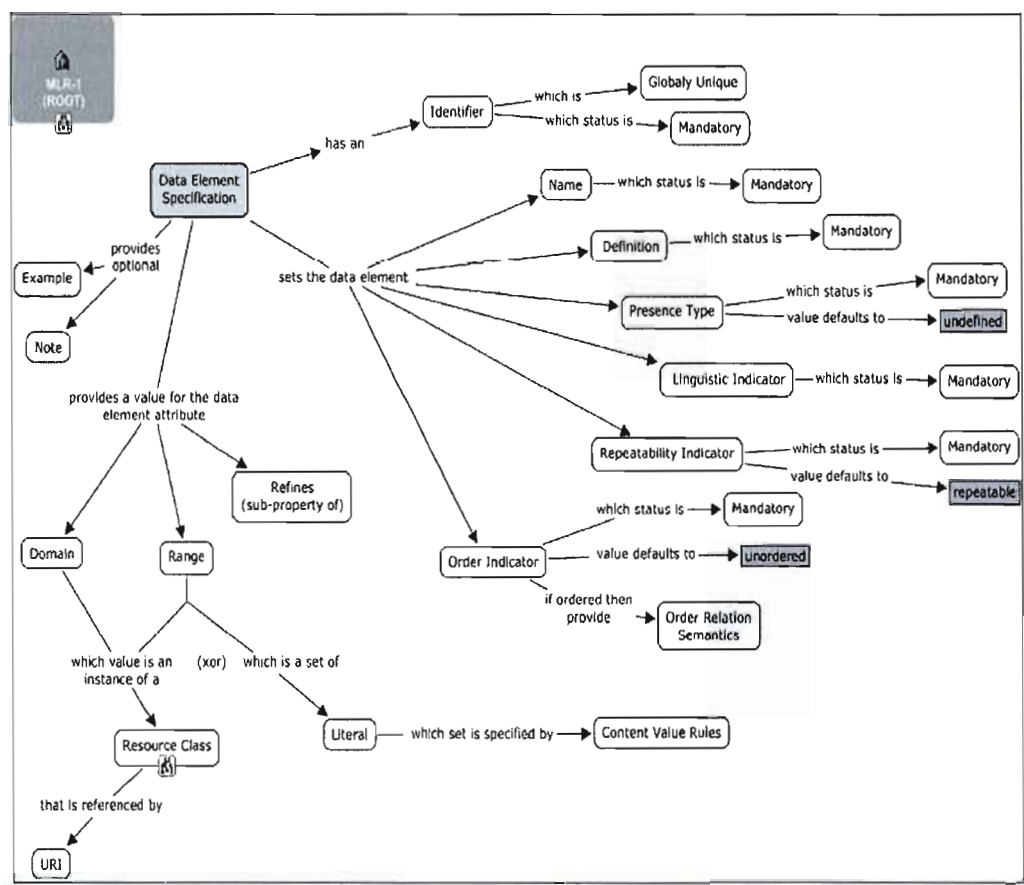


Figure 4.2 : Carte conceptuelle d'une spécification d'un élément de métadonnées (WG4, 2009)

Toutefois, cette partie ne spécifie pas les règles relatives à l'enregistrement des notices de métadonnées MLR. Une partie importante du document est consacrée aux attributs courants qui sont associés à chaque élément de métadonnée (figure 4.3).

- **Identifier** (data element specification identifier)
- **Name** (data element name)
- **Definition** (data element definition)
- **Presence type** (data element presence type)
- **Linguistic indicator** (data element linguistic indicator)
- **Repeatability indicator** (data element repeatability indicator)
- **Order indicator** (data element order indicator)
- **Order relation description**
- **Domain** (data element domain)
- **Range** (data element range)
- **Content value rules**
- **Refines**
- **Example(s)**
- **Note(s)**

Figure 4.3 : Attributs communs aux éléments MLR (WG4, 2009)

Certains attributs d'éléments de métadonnées méritent une attention particulière. L'*identifiant* de la spécification d'un élément de données est une valeur non ambiguë, globalement unique et linguistiquement neutre, attribuée par une autorité reconnue. Le *nom* d'un élément doit être unique à travers les différentes parties de la norme. L'attribut «range» définit l'étendue ou le domaine des valeurs qu'un élément de métadonnées MLR peut prendre. Dans le cas où la valeur de l'attribut est un littéral, la norme recommande de spécifier un **espace de valeurs** (value space) et / ou un espace lexical (lexical space). L'espace des valeurs contient l'ensemble des valeurs admissibles que peut prendre un élément de métadonnée.

L'**espace lexical** correspond, quant à lui, à la représentation visuelle d'une valeur. Ainsi, la valeur 4 peut être représentée de diverses façons : quatre, "4", "iv", etc. Mais il faut choisir une et une seule **représentation canonique** basée sur des **règles de représentation canonique** bien définies (figure 4.4).

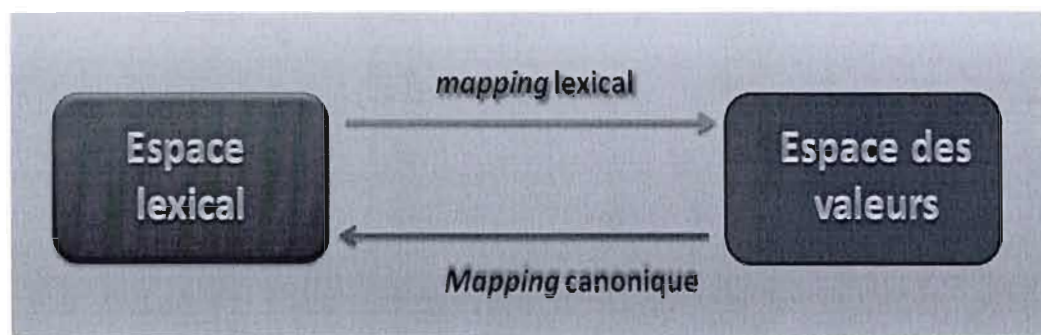


Figure 4.4 : Mapping canonique et lexical (traduit de ISO/CEI JTC1 SC36, 2008c)

Ces définitions sont pour le moment provisoires mais l'exemple suivant permet de clarifier le sens de ces termes (ISO/CEI JTC1 SC36, 2008c).

Espace des valeurs	{T, F}
Espace lexical	{"0", "1", "true", "false"}
Mapping lexical	{<"true", T>, <"1", T>, <"0", F>, <"false", F>}
Mapping canonique	{<T, "true">, <F, "false">}

La représentation canonique vise à l'internationalisation et l'interopérabilité sémantique. Par ailleurs, à chaque élément, un certain nombre d'indicateurs peuvent être associés (figure 4.5).

Indicateur de présence	<ul style="list-style-type: none"><li>• Conditionnel</li><li>• Obligatoire</li><li>• Optionnel</li></ul>
Indicateur linguistique	<ul style="list-style-type: none"><li>• Oui</li><li>• Non</li></ul>
Indicateur de répétitions	<ul style="list-style-type: none"><li>• Oui</li><li>• Non</li></ul>
Indicateur d'ordre	<ul style="list-style-type: none"><li>• Oui</li><li>• Non</li><li>• Non applicable</li></ul>

Figure 4.5 : Indicateurs des caractéristiques d'un élément de métadonnée

Trois concepts importants sont décrits dans la partie 1. Outre le concept de ressource d'enseignement et d'apprentissage, l'élément de métadonnée est la base de la description d'une ressource. L'ensemble des éléments utilisés pour la description d'une ressource donnée constitue une notice MLR. Celle-ci est définie dans la norme comme étant un ensemble d'éléments de métadonnées décrivant

une ressource d'enseignement et d'apprentissage ainsi que certaines autres entités reliées à celle-ci.

Pour favoriser l'adoption du MLR à travers le monde, la partie 1 énonce les principes qui doivent guider le développement des différentes parties de la norme :

- prévoir des mécanismes d'extension afin de mieux répondre aux besoins de communautés aux horizons culturels et linguistiques différents;
- maximiser l'utilisation des normes et standards existants;
- favoriser la compatibilité avec le LOM et le Dublin Core et faciliter ainsi leur migration complète vers le MLR;
- maximiser l'interopérabilité sémantique;
- supporter plusieurs niveaux de granularité dans la description des ressources;
- prévoir des règles précises permettant d'assurer l'interopérabilité des profils d'application du MLR, etc.

La partie 2 du MLR est une application des structures de données et des règles définies dans la première partie de la norme. Elle spécifie un ensemble d'éléments de base permettant de décrire de façon minimale une ressource d'enseignement et d'apprentissage.

#### ***4.2.2. Les éléments de base du MLR***

La partie 2 du MLR fournit une liste d'éléments de base. Certains ont pour but de décrire une ressource ; d'autres ont plutôt un rôle fonctionnel dans la mesure où ils ont été conçus pour supporter l'interopérabilité nécessaire à l'échange des notices de métadonnées entre différents dépôts numériques. La partie 2 se limite à un

noyau d'éléments considérés comme essentiels. En tout, 19 éléments de métadonnées ont été identifiés (Annexe 1).

ISO\_IEC\_19788-2:2008::DES0080 Pedagogical Type

Attribute	Attribute value (string)
Identifiant	ISO_IEC_19788-2:2008::DES0080
Name	{{(Pedagogical Type, eng), (Approche pédagogique, fra)}}
Definition	{{(recommended means, process or application through which the learning resource contributes to teaching or learning, eng), (utilisation recommandée de la ressource pour la transmission des connaissances, attitudes ou habiletés, fra)}}
Presence type	3 (optional)
Structure indicator	1 (stand-alone)
Linguistic indicator	2 (non-linguistic)
Repeatability indicator	10 (any number of occurrences)
Ordered indicator	2 (not significant)
Order relation	{}
Lexical space	[[VOCABULARY]] (see ISO_IEC_19788-2:2008::VDES0080 in section 8)
Canonical representation rules	PRS0012 (see section 7.2.3)
Example	ISO_IEC_19788-2:2008::VDES0080:T01
Comment	-

Figure 4.6 : Exemple d'un élément MLR et ses attributs

ISO\_IEC\_19788-2:2008::VDES0080, Pedagogical Type

ID Code	eng (ISO) Pedagogy Type G Term	Description
(1)	(2) (3)	(4)
01	09 Learning Measurement	measurement of a learner's knowledge or understanding
02	09 Problem Solving	analysis of a particular situation or environment and expression of a solution
03	09 Learning Sequence	organized sequence of learning tasks
04	09 Tool	instrument to complete a specific task in a larger process
05	09 Presentation	predominantly visual and/or oral presentation
06	09 Description	resource used to give descriptive information
07	09 Explanation	clarification of a situation or phenomenon
08	09 Exercise	a task performed or problem solved in order to develop skill or understanding
09	09 Experiment	conduction of a set of procedures and observations usually involving data collection
10	09 Simulation	modeled representation of a situation or phenomenon

Figure 4.7 : Exemple d'un vocabulaire MLR



Certains éléments de métadonnées utilisent un domaine de valeurs codifiées ou des vocabulaires pour être renseignés (figure 4.6). La figure 4.7 montre un vocabulaire utilisé pour renseigner l'élément de métadonnées « *Pedagogical Type* ».

La partie 1 du MLR fournit une approche systématique pour spécifier un ensemble commun d'attributs qui doivent s'appliquer à chaque élément de métadonnées. Le respect des règles définies dans la partie 1 assure la précision du sens de chaque élément et garantit l'interopérabilité sémantique. De plus, les différentes caractéristiques du MLR font de lui une norme permettant de décrire les ressources d'enseignement et d'apprentissage en tenant compte des exigences du Web sémantique. C'est ce que nous verrons en détail au prochain chapitre.



## CHAPITRE V

### LE MLR ET LE WEB SÉMANTIQUE

Les concepteurs du MLR avaient de grands défis à relever : assurer l'interopérabilité des métadonnées dans un contexte multiculturel et multilingue et répondre aux exigences du Web sémantique tout en garantissant la compatibilité avec le LOM mais sans hériter de ses faiblesses. Sur ce dernier point, les inquiétudes étaient nombreuses.

#### 5.1. Atouts du MLR pour le Web sémantique

Le modèle abstrait du MLR n'est pas encore disponible. Toutefois, l'approche centrée sur la *ressource* qui a été adoptée par concepteurs du MLR laisse penser que son modèle abstrait ressemblera à celui du Dublin Core et sera tout à fait compatible avec le modèle RDF (figures 5.1).

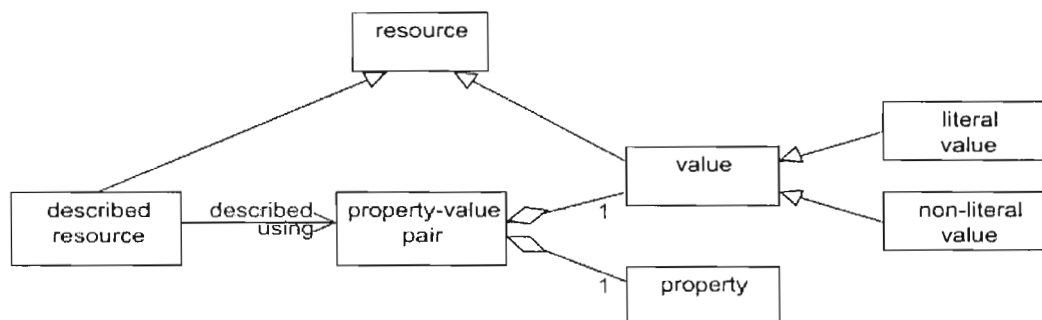
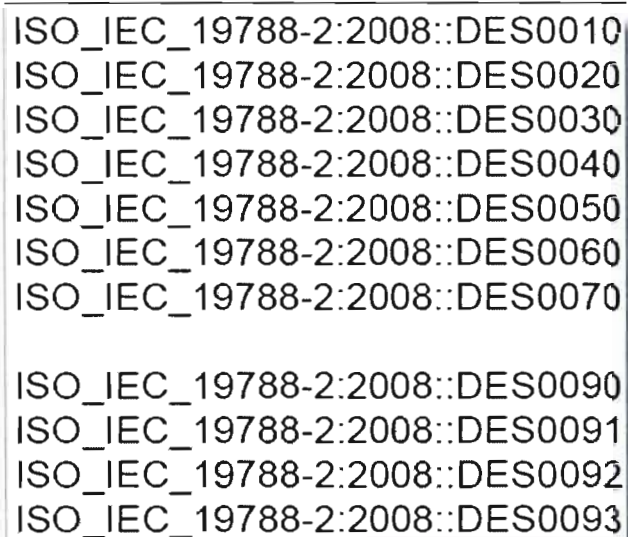


Figure 5.1 : Le modèle abstrait du Dublin Core (Nilsson et al., 2007)

De cette constatation, on peut conclure que l'approche RDF, qui semble privilégiée, lui assure la conformité avec les exigences du Web sémantique.

La prise en compte du contexte socioculturel et linguistique dans lequel les ressources sont créées et utilisées est un apport important du MLR par rapport au LOM. Dans le LOM, le nom d'un élément de métadonnées est porteur de sa signification. L'argument selon lequel les noms d'éléments ne doivent pas être considérés comme des mots dans une langue particulière mais comme des chaînes de caractères alphanumériques qui identifient un élément de métadonnée de manière non ambiguë et de ce fait, interopérable, est peu convaincant. L'approche choisie par le MLR résout ce problème. L'utilisation d'un identifiant linguistiquement neutre (figure 5.2), unique et sans ambiguïté, permet aux utilisateurs de donner la même signification à chaque élément sans se focaliser sur son nom.



```
ISO_IEC_19788-2:2008::DES0010
ISO_IEC_19788-2:2008::DES0020
ISO_IEC_19788-2:2008::DES0030
ISO_IEC_19788-2:2008::DES0040
ISO_IEC_19788-2:2008::DES0050
ISO_IEC_19788-2:2008::DES0060
ISO_IEC_19788-2:2008::DES0070

ISO_IEC_19788-2:2008::DES0090
ISO_IEC_19788-2:2008::DES0091
ISO_IEC_19788-2:2008::DES0092
ISO_IEC_19788-2:2008::DES0093
```

Figure 5.2 : Exemple d'identifiants linguistiquement neutres

```

<?xml version="1.1" encoding="UTF-8" ?>
<mlr xmlns="http://www.iso.org/jtc1/sc36/MLR"
  xmlns:part2="urn:ISO_IEC_19788-2:2008::">

  <part2:des0020>
    <string language="fra">Créer•Réutiliser•Diffuser•Partager</string>
    <string language="eng">Create•Reuse•Distribute•Share</string>
    <string language="cmn">创造•重用•分发•共享</string></title>
  </part2:des0020>

  <part2:des0090>
    <part2:des0091>text/html</part2:des0091>
    <part2:des0093>http://www.uquebec.ca/reauq-pi/</part2:des0093>
  </part2:des0090>

  <part2:des0090>
    <part2:des0091>application/pdf</part2:des0091>
    <part2:des0093>http://www.uquebec.ca/reauq-pi/pdf.php?id=9</part2:des0093>
  </part2:des0090>

  <part2:des0110>
    <part2:des0111>
      <string language="fra">Paternité-Pas d'utilisation commerciale-Pas de modification 2.5
        Canada
      </string>
      <string language="eng">Attribution-Noncommercial-No Derivative Works 2.5
        Canada
      </string>
      <string language="zho">署名-非商业性使用-禁止演绎 2.5 加拿大</string>
    </part2:des0111>
    <part2:des0112>
      <string language="eng">http://creativecommons.org/licenses/by-nc-nd/2.5/ca/deed.en_CA
      </string>
      <string language="zho">http://creativecommons.org/licenses/by-nc-nd/2.5/ca/deed.zh
      </string>
    </part2:des0112>
  </part2:des0110>

</mlr>

```

Figure 5.3 : Représentation des éléments MLR dans la machine

Lorsque la signification d'un élément est bien définie et partagée, il est possible pour les utilisateurs de lui donner plusieurs noms selon leurs langues et leurs cultures (<sup>67</sup>).

<sup>67</sup> Ce processus est appelé "régionalisation", c'est-à-dire le fait de traduire le contenu d'un document, un concept, une interface logicielle dans une autre langue en tenant compte de la culture locale. Le processus inverse est appelé « internationalisation » et vise à rendre un produit, un document, une

Une telle approche résout le problème du multilinguisme. Ainsi, pour un élément de métadonnées, les équivalents linguistiques sont utilisés pour l'interface avec l'utilisateur humain (figure 5.4); mais derrière ces mots, il doit y avoir la valeur de l'identifiant telle que fournie dans le MLR (figure 5.3).

```
<?xml version="1.1" encoding="UTF-8" ?>
<mlr xmlns="http://www.iso.org/jtc1/sc36/wg4/MLR">
  <title>
    <string language="fra">Créer*Réutiliser*Diffuser*Partager</string>
    <string language="eng">Create*Reuse*Distribute*Share</string>
    <string language="cmn">创建*重用*分发*共享</string></title>
  </title>
  <resourceInstance>
    <format>text/html</format>
    <location>http://www.quebec.ca/reaug-pi/</location>
  </resourceInstance>
  <resourceInstance>
    <format>application/pdf</format>
    <location>http://www.quebec.ca/reaug-pi/pdf.php?id=9</location>
  </resourceInstance>
  <rights>
    <rightsDescription>
      <string language="fra">Paternité-Pas d'utilisation commerciale-Pas de modification 2.5
        Canada
      </string>
      <string language="eng">Attribution-Noncommercial-No Derivative Works 2.5
        Canada
      </string>
      <string language="zho">署名-非商业性使用-禁止演绎2.5 加拿大</string>
    </rightsDescription>
    <rightsSource>
      <string language="eng">http://creativecommons.org/licenses/by-nc-nd/2.5/ca/deed.en_CA
      </string>
      <string language="zho">http://creativecommons.org/licenses/by-nc-nd/2.5/ca/deed.zh
      </string>
    </rightsSource>
  </rights>
</mlr>
```

Figure 5.4 : Les éléments MLR pour un utilisateur humain anglophone

application facilement adaptable à différentes langues et cultures. On rencontre souvent les abréviations i18n et i10n pour désigner les deux termes, « *internationalization* » et « *localization* » parce qu'il y a 18 caractères entre le i et le n du premier et 10 lettres pour le dernier.

L'approche préconisée par le MLR apporte donc une contribution significative à l'interopérabilité des métadonnées. Elle répond à une nouvelle vision des métadonnées selon laquelle celles-ci sont d'abord destinées à la machine (Nilsson, 2008). Nous verrons, dans la section ci-dessous, une façon d'intégrer des éléments MLR dans des pages XHTML et des les récupérer ensuite sous forme de triplets RDF pour permettre à la fois aux machines et aux humains d'accéder facilement aux ressources d'enseignement et d'apprentissage disponibles sur le Web (Feigenbaum & Torres, 2006).

## **5.2. Intégration des éléments de métadonnées *MLR* dans le Web sémantique**

Nous avons évoqué au chapitre 4 l'intérêt d'indexer les ressources d'enseignement et d'apprentissage avec les métadonnées normalisées afin d'en faciliter le repérage et la mutualisation. Cette normalisation « assure que les utilisateurs (et les logiciels) partagent non seulement le nom des éléments de métadonnées mais aussi la façon de les renseigner » (Bourda, 2008, p. 24) et facilite ainsi le partage, l'échange et la réutilisation des descriptions de ressources dans le but, notamment, de réduire les coûts reliés au processus d'indexation. L'efficacité des systèmes de repérage de l'information repose, en partie, sur la richesse sémantique des schémas de métadonnées adaptés aux besoins pratiques de diverses communautés. Aussi, le MLR constitue-t-il un référentiel précieux mais son utilisation à l'échelle du Web sémantique pourrait se heurter à de défis de taille compte tenu de la complexité des technologies actuellement disponibles. Il nous paraît donc important de trouver des moyens relativement simples permettant de distribuer des éléments de métadonnées MLR à travers le Web sémantique. Avant de présenter l'approche que nous préconisons, il convient de décrire les différents contextes d'utilisation du MLR (figure 5.5).

### 5.2.1. Contextes d'utilisation du MLR

La situation la plus courante, est celle où une institution d'enseignement développe pour ses besoins propres un dépôt de ressources d'enseignement et d'apprentissage et adopte le MLR comme langage d'indexation de ces ressources (modèle centralisé). Mais, les ressources d'enseignement et d'apprentissage peuvent aussi être localisées en dehors des dépôts d'objets d'apprentissage et particulièrement sur le Web (modèle distribué). La différence entre les deux est que dans le cas du Web, on a une masse de REA généralement sans aucune indexation. Par contre, les banques ou dépôts d'objets d'apprentissage disposent généralement d'un référentiel qui regroupe les métadonnées relatives à ces ressources. Bref, on peut distinguer deux types de répertoires : d'une part, ceux qui contiennent les REA et les métadonnées qui les décrivent et, d'autre part, ceux qui abritent exclusivement des métadonnées des REA éparpillées dans divers réseaux.

		Ressources	
Métadonnées		centralisées	distribuées
	centralisées	✓ dépôt de REA + référentiel	✓ référentiel notices MLR
	distribuées	✓	✓ ressources html avec balises <i>meta</i>

Figure 5.5 : Contextes d'utilisation du MLR



Cette dichotomie ne s'applique pas qu'aux ressources mais aussi aux métadonnées pour lesquelles on peut avoir un modèle de déploiement **centralisé** ou **distribué**. Le premier fait penser au catalogue de bibliothèques géré par une base de données. Le deuxième peut être envisagé de deux manières (figure 5.6). On peut envisager, par exemple, le cas de plusieurs universités disposant chacune de son dépôt de REA avec leurs métadonnées. Le modèle distribué repose dans ce cas sur la mise en réseau des différents dépôts institutionnels. Les métadonnées décrivant les ressources sont alors déployées sur des serveurs différents. Ce modèle représente certainement un avantage évident puisqu'il permet aux utilisateurs d'accéder à un référentiel plus important, mais le modèle distribué basé sur le Web est le plus intéressant.



Figure 5.6 : Métadonnées centralisées et distribuées

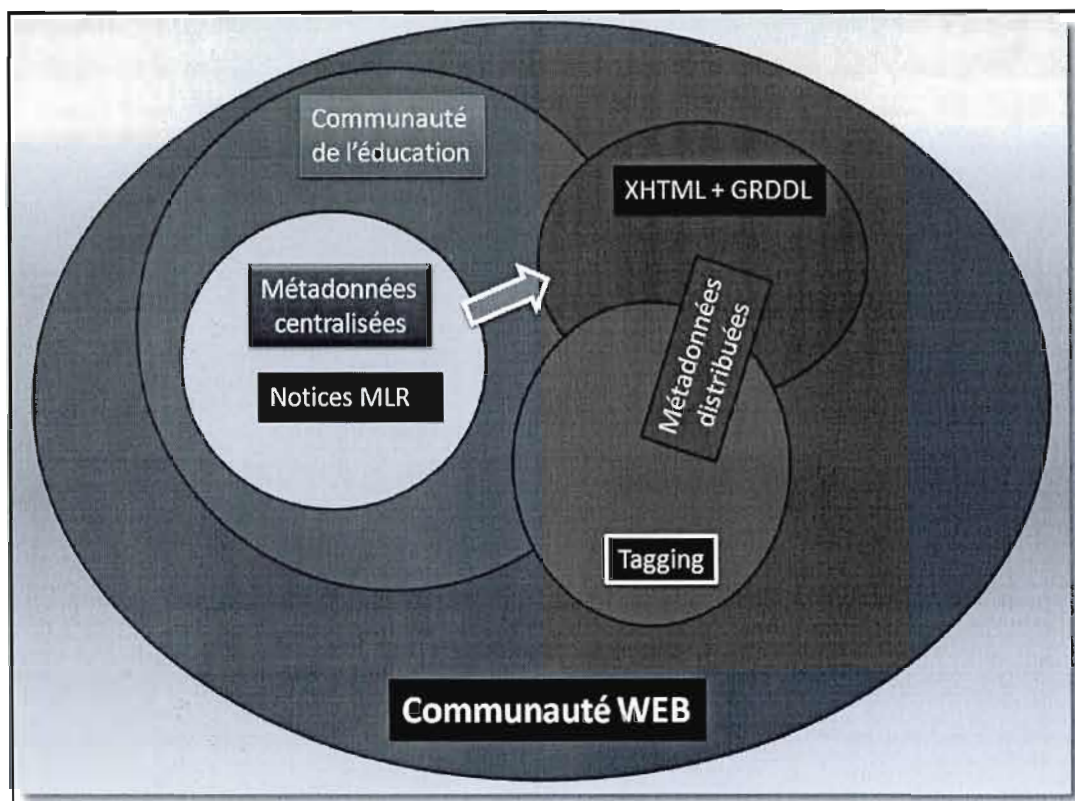


Figure 5.7 : Distribution des métadonnées MLR dans le Web

Les métadonnées distribuées sur le Web ont non seulement une plus grande visibilité mais aussi elles peuvent être réutilisées pour décrire d'autres ressources. Les URIs assurent qu'un élément de métadonnées soit associé à une définition unique qui peut être utilisée sans aucune ambiguïté de sens. Les espaces de nommage permettent de pointer vers un fichier qui contient la définition exacte d'un élément de métadonnées. L'intégration des éléments MLR dans les pages Web est facilitée grâce à ses qualités intrinsèques, particulièrement l'utilisation des URI comme moyen d'identification à la fois des ressources décrites, des spécifications d'éléments de métadonnées et de la valeur des attributs à partir des vocabulaires. L'assignation d'une URI à un élément ou à la valeur de cet élément signifie qu'il peut être cité, réutilisé de façon non ambiguë (Nilsson, 2008).



Le problème qu'il faut résoudre est illustré à la figure 5.7. Comment distribuer sur le Web des métadonnées centralisées dans un référentiel particulier et qui portent sur tout type de ressource (page HTML, document PDF, document vidéo ou iconographique, etc.) ? Plus précisément, comment rendre disponibles les éléments de métadonnées MLR aux applications du Web sémantique ? En outre, comment exploiter le MLR dans le contexte du Web 2.0 où l'indexation collaborative ou distribuée est devenue incontournable ? Nous allons donc considérer deux cas, à savoir : 1° la distribution des éléments ou de notices de métadonnées MLR dans le Web sémantique, 2° la contribution distribuée à l'indexation, communément appelée indexation collaborative.

### ***5.2.2 Approche de distribution des notices de métadonnées MLR***

Au chapitre 3, nous avons abordé la notion de microformats. Nous avons vu que ces microformats offraient une voie pragmatique vers le Web sémantique (Khare & Celik, 2006). Malheureusement, les microformats sont orientés vers des applications spécifiques et n'offrent pas l'extensibilité essentielle à l'interopérabilité (Adida, 2008). Cette solution ne peut donc être retenue pour le MLR. Les attributs RDF (ou RDFa) offrent donc la solution la plus efficace. Cette solution nous amène à distinguer deux cas.

#### **Cas 1 : ressource en format (X)HTML**

Le cas le plus simple est celui où on a une ressource (X)HTML. Il suffit de mettre les attributs RDFa dans le *HEAD* du document (figure 5.8).

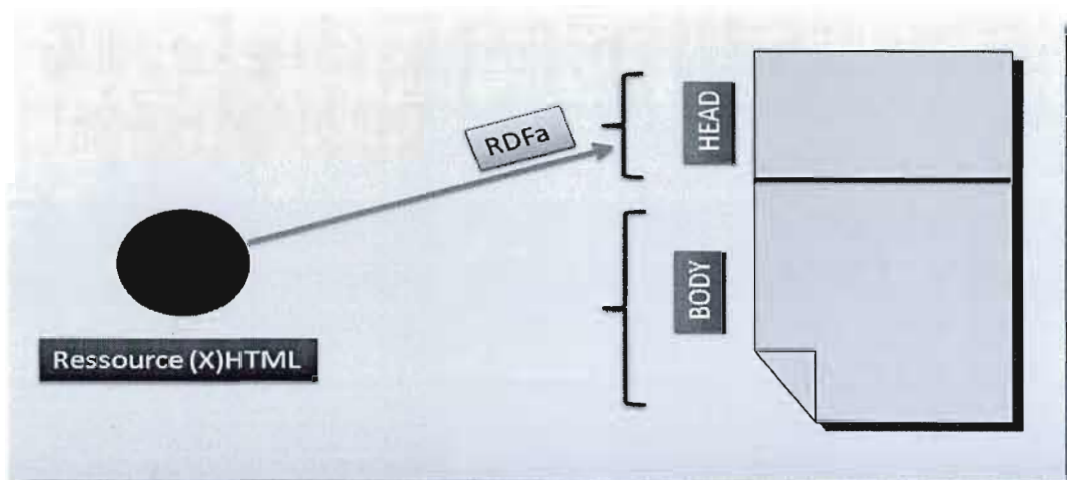


Figure 5.8 : Ressource en format (X)HTML et attributs RDF

### Cas 2 : Ressource quelconque

Le deuxième cas est celui où on a une ressource quelconque à laquelle correspond une notice MLR conservée dans une base de données à la manière d'un catalogue de bibliothèque.

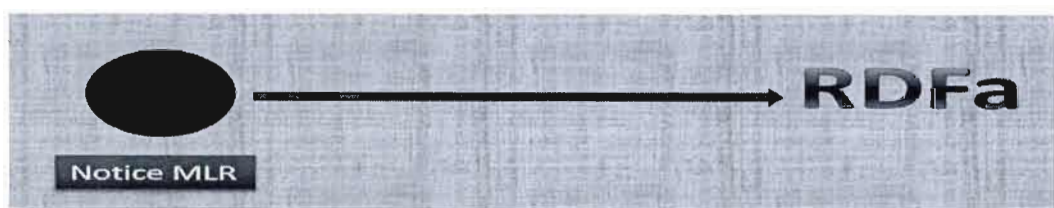


Figure 5.9 : Notice MLR vers RDFa

La solution consiste à convertir, à l'aide d'une feuille de style XSLST, tous les éléments de la notice MLR en attributs RDFa (figures 5.9 et 5.10). Il faut alors construire un « container » en (X)HTML qui va accueillir, dans le *HEAD*, les éléments de métadonnées MLR convertis en attributs RDF. L'élément *<BODY>* est vide.

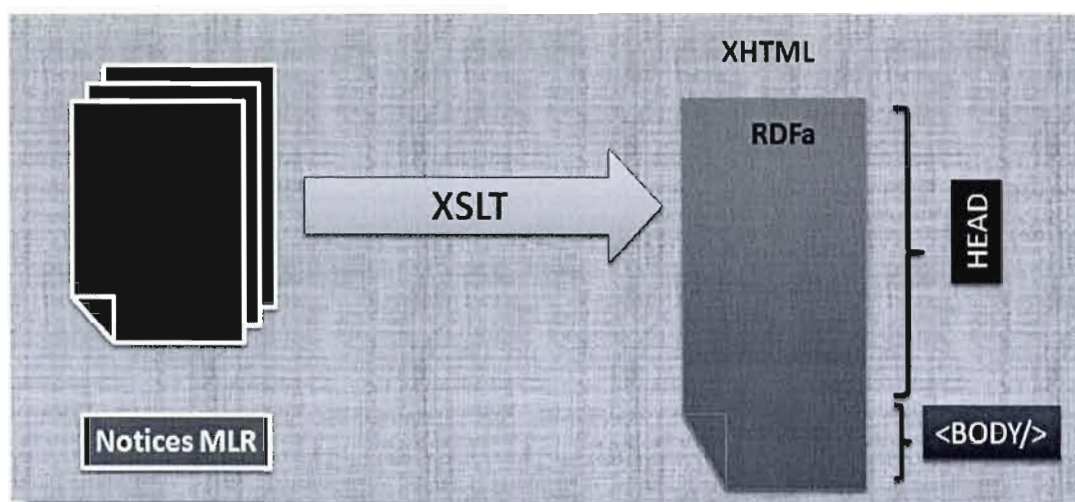


Figure 5.10 : Conversion d'une notice MLR avec XSLT en attributs RDF

En prenant comme exemple une courte notice MLR reprise ci-dessous, voici, à quoi ressemblerait une page XHTML contenant les attributs RDFa. Une fois que le mapping XML pour le MLR sera défini, il sera relativement facile d'écrire une feuille de style permettant de convertir une notice MLR en attributs RDFa.

```
<?XML version="1.1" encoding="UTF-8" ?>
<mlr xmlns="http://www.iso.org/jtc1/sc36/wg4/MLR">
  <learningResourceIdentifier>Créer, réutiliser, diffuser, partager</
learningResourceIdentifier >
  <title>Créer, réutiliser, diffuser, partager</title>
  <language>fra</language>
  <lastUpdate>2008-12-12</lastUpdate>
  <contributionPersons>Dupont Jean</contributionPerson>
  <contributionRole>éditeur</contributionRole>
</mlr>
```

```

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN" "http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
<html xmlns:mlr="http://www.iso.org/jtc1/sc36/wg4/MLR">

<head profile="http://ns.inria.fr/grddl/rdfa/">
<meta about="http://example.com/rea/123">
  <meta property="mlr:title">Le MLR</meta>
  <meta property="mlr:language">fra</meta>
  <meta property="mlr:lastUpdate">2006-10-01</meta>
  <meta property="mlr:contributionPerson">Jean Dupont</meta>
  <meta property="mlr:contributionRole">éditeur</meta>
</meta>
</head>
<body/>
</html>

```

Plusieurs outils permettant de reconnaître des balises RDFa dans une page HTML sont disponibles (voir sur le site <http://esw.w3.org/topic/CustomRdfDialects>). Un outil développé par F. Gandon (2008) est particulièrement intéressant. Pour l'utiliser, il suffit d'indiquer dans l'entête de la page XHTML le profil indiquant que cette page contient du RDF/A. Cette indication est mise dans l'entête de la page. Ensuite, on indique également dans l'entête, la feuille de style qui fera la transformation comme le montre le code ci-dessous :

```

<head profile="http://www.w3.org/2003/g/data-view">

(...)
<linkrel="transformation"
href="http://ns.inria.fr/grddl/rdfa/2008/09/03/RDFa2RDFXML.xsl"/>
(...)
</head>.

```

Par la suite, il faut recourir à un service GRDDL comme celui fourni par le W3C pour faire l'extraction (figure 5.11).

**W3C** **GRDDL Service**  
Extracting RDF from XHTML/XML using **GRDDL**

Validate by URI    Validate by Direct Input

Validate by direct input

Enter the document markup to GRDDL:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN" "http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
<html xmlns:mlr="http://www.iso.org/jtc1/sc36/wg4/MLR">

<head profile="http://ns.inria.fr/grddl/rdfa/">
  <meta about="http://example.com/rea/123">
    <meta property="mlr:title">Le MLR</meta>
    <meta property="mlr:language">fra</meta>
    <meta property="mlr:lastUpdate">2006-10-01</meta>
    <meta property="mlr:contributionPerson">Jean Dupont</meta>
    <meta property="mlr:contributionRole">éditeur</meta>
  </meta>
</head>
<body/>
</html>
```

Output format: RDF/XML as application/rdf+xml

get GRDDL results

Figure 5.11 : Service GRDDL du W3C (Hazaël-Massieux, 2008)

Ce service GRDDL peut prendre comme paramètre l'adresse de la page XHTML qui contient les attributs RDFa ou le texte complet d'une page HTML. Il permet de choisir le format de sortie approprié en fonction des besoins de l'application. Le résultat de cette opération est illustré à la figure 5.12. La figure 5.13 résume le processus mis en œuvre dans cette opération.

```

<?xml version="1.0" encoding="utf-8" ?>
- <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
-   <rdf:Description rdf:about="http://example.org/">
-     <ns0:MLRtitle xmlns:ns0="http://www.iso.org/jtc1/sc36/wg4/">Le MLR</ns0:MLRtitle>
-   </rdf:Description>
-   <rdf:Description rdf:about="http://example.org/">
-     <ns0:MLRlanguage xmlns:ns0="http://www.iso.org/jtc1/sc36/wg4/">fra</ns0:MLRlanguage>
-   </rdf:Description>
-   <rdf:Description rdf:about="http://example.org/">
-     <ns0:MLRlastUpdate xmlns:ns0="http://www.iso.org/jtc1/sc36/wg4/">2006-10-01</ns0:MLRlastUpdate>
-   </rdf:Description>
-   <rdf:Description rdf:about="http://example.org/">
-     <ns0:MLRcontributionPerson xmlns:ns0="http://www.iso.org/jtc1/sc36/wg4/">Jean Dupont</ns0:MLRcontributionPerson>
-   </rdf:Description>
-   <rdf:Description rdf:about="http://example.org/">
-     <ns0:MLRcontributionRole xmlns:ns0="http://www.iso.org/jtc1/sc36/wg4/">éditeur</ns0:MLRcontributionRole>
-   </rdf:Description>
- </rdf:RDF>
<!-- Extracted from pasted markup by http://www.w3.org/2007/08/grddl/ -->

```

Figure 5.12 : Résultat d'une transformation GRDDL

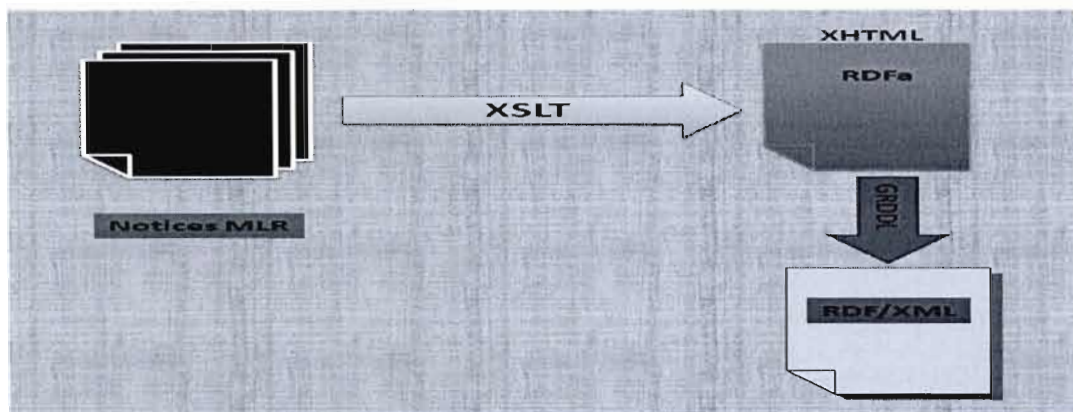


Figure 5.13 : Processus de conversion d'une notice MLR en RDFa, puis en RDF/XML

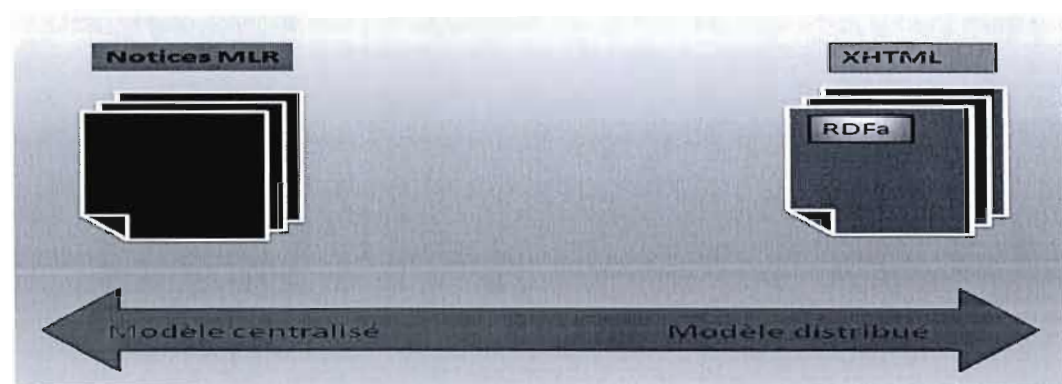


Figure 5.14 : Modèle distribué et centralisé des métadonnées

Le RDFa permet de distribuer les éléments de métadonnées MLR dans le Web (figure 5.14). Le travail d'indexation des ressources est un autre défi à relever. Compte tenu de sa complexité et de la quantité des ressources qu'il contient, le Web d'aujourd'hui pose des défis insurmontables aux techniques traditionnelles d'indexation des ressources d'information. Aussi, une nouvelle technique d'indexation collective est-elle venue à la rescousse de ces pratiques : le *tagging*. De nombreuses ressources d'enseignement et d'apprentissage sont disponibles et une grande partie n'est pas indexée. Comment le tagging peut-il contribuer à la distribution des métadonnées MLR dans le Web ?

### **5.2.3 Contributions distribuées à l'indexation et le MLR**

L'indexation des ressources est une tâche complexe et coûteuse. Dans le nouveau paysage informationnel, l'intelligence humaine est exploitée pour indexer collectivement une masse de ressources sans cesse croissante. Cette approche collaborative utilise les capacités sociales des utilisateurs pour « filtrer, indexer, recommander l'information et échanger les métadonnées sur les ressources à travers des réseaux sociaux » (Ta Tuan, 2005, p. 40) tels que Delicio.us <sup>(68)</sup>, Flickr <sup>(69)</sup>, LibraryThing <sup>(70)</sup>, etc. Ces métadonnées sont données par les utilisateurs eux-mêmes et sont partagées dans une sorte de mémoire collective pour faciliter l'accès à l'information (Ta Tuan, 2005, p. 5). Ces *tags* sont des mots-clés ajoutés aux ressources par des utilisateurs à des fins de description, de localisation,

---

<sup>68</sup> <http://delicious.com/>

<sup>69</sup> <http://flickr.com>

<sup>70</sup> <http://www.librarything.com>



d'évaluation, de commentaires, etc. (Smith, 2008). Simples, flexibles, extensibles et faciles d'utilisation, les tags améliorent la recherche de l'information. Contrairement à la technique d'indexation traditionnelle qui repose sur les vocabulaires contrôlés, le tagging permet à chacun d'utiliser son propre vocabulaire ou une folksonomie <sup>(71)</sup>.

Les deux approches présentent chacune ses avantages et inconvénients. L'utilisation d'un vocabulaire commun minimise l'hétérogénéité dans l'indexation mais, en même temps, requiert des efforts importants dans la construction et le maintien du vocabulaire. De plus, les vocabulaires contrôlés sont peu flexibles et évoluent très lentement <sup>(72)</sup>. En contrepartie, dans les folksonomies, il n'y a pas de relations sémantiques entre les différents tags comme dans les vocabulaires traditionnels. Ces relations sont plutôt construites sur base d'inférence statistique (Smith, 2008, p. 86). Malgré des avantages indéniables, les folksonomies n'ont pas la prétention de se substituer aux outils traditionnels d'indexation (thesaurus, listes de vedettes matière, etc.). Barbosa (2008) préconise une approche hybride qui permet de combiner les avantages des vocabulaires contrôlés et des folksonomies.

Cette approche présente un intérêt particulier pour le MLR. Imaginons un référentiel de ressources d'enseignement et d'apprentissage qui serait conçu selon le modèle de LibraryThing <sup>(73)</sup> et où les utilisateurs seraient appelés à signaler les ressources

---

<sup>71</sup> Un ensemble des tags constitue une folksonomie.

<sup>72</sup> À titre d'exemple, le descripteur « apartheid » n'a été ajouté au répertoire des vedettes matières de la Library of Congress qu'en 1986 ! (Smith, 2008, p. 87)

<sup>73</sup> *LibraryThing* est un site Web où plus de 200 000 amateurs de livres à travers le monde décrivent, à l'aide des tags, les livres de leurs collections personnelles. Le site recense actuellement plus de 15 millions de livres, faisant de lui, le deuxième catalogue de livres en Amérique du Nord (Smith, 2008, p. 2).



dans le référentiel et à leur associer des tags (figure 5.15). Tout comme dans *LibraryThing*, le but visé est d'avoir un nombre aussi élevé que possible de participants actifs prêts à contribuer à la description des ressources qui les intéressent. Pour le monde de l'éducation, un tel projet permettrait de construire une base importante de ressources pédagogiques. Il faciliterait la collaboration des utilisateurs par l'ajout ou le signalement des ressources et l'obtention des métadonnées descriptives des ressources partagées à un coût très faible par rapport à l'approche traditionnelle d'indexation qui est trop coûteuse malgré ses qualités indéniables. On peut citer en exemple le projet **MERLOT** <sup>(74)</sup>, un répertoire américain de ressources d'enseignement et d'apprentissage en ligne destiné à l'enseignement supérieur.

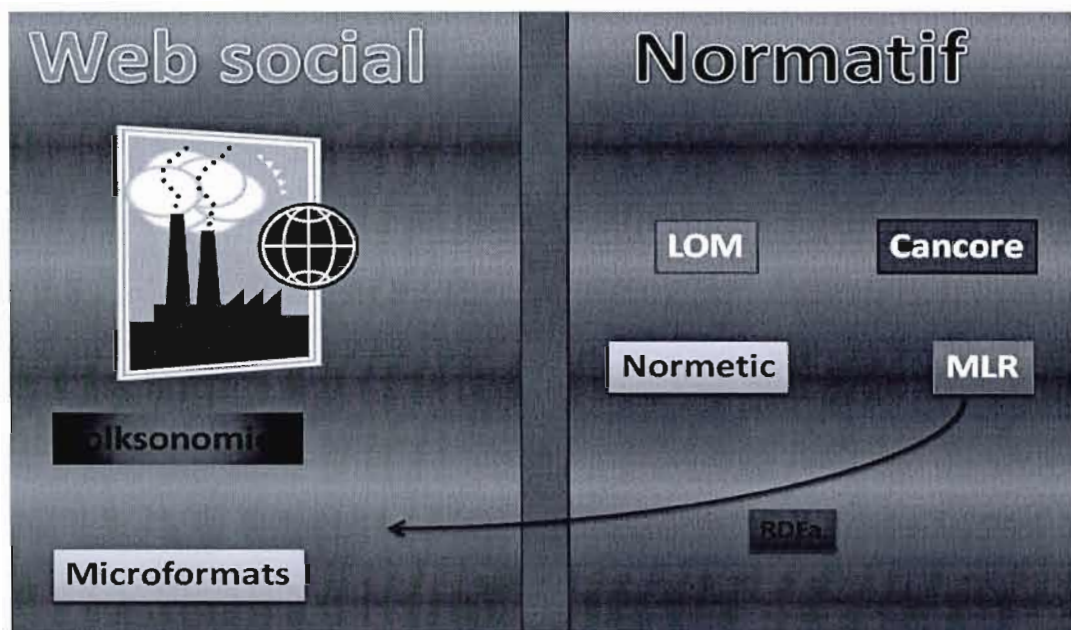


Figure 5.15 : Le MLR et le tagging

<sup>74</sup> Multimedia Educational Resource for Learning and Online Teaching

De façon générale, le système de *tagging* donne aux utilisateurs la liberté d'utiliser les termes de leur choix. Cependant, si dans des applications personnelles on peut laisser libre court à l'imagination de chacun, dans les systèmes à caractère social ou professionnel, il est souhaitable d'exercer un certain contrôle; l'idée étant de permettre à l'utilisateur d'insérer des tags de son choix mais aussi de le guider dans ce processus pour favoriser l'usage des métadonnées normalisées.

Une interface appropriée permet de guider l'utilisateur tout en lui donnant la liberté nécessaire pour la saisie des tags issus de son vocabulaire. Cela peut s'avérer pratique pour un portail destiné aux ressources d'enseignement et d'apprentissage. L'utilisateur peut utiliser les tags de son choix, mais le formulaire permettra également d'avoir un minimum de métadonnées communes. On peut donc imaginer un référentiel de ressources d'enseignement et d'apprentissage qui utiliserait le MLR comme schéma de descriptions des ressources signalées par les utilisateurs eux-mêmes. Les éléments de métadonnées MLR n'étant pas obligatoires, on peut offrir à l'utilisateur un formulaire où seuls les éléments jugés essentiels seraient complétés pour chaque ressource signalée : par exemple le titre, la langue, la description, l'approche pédagogique, la contribution à la ressource, le rôle joué dans la création de la ressource, la date de participation à la ressource, le format, la taille, la localisation, les droits, etc. Une notice MLR peut alors être constituée à partir des informations saisies par les contributeurs. Le mécanisme de conversion vu à la section précédente permettrait de distribuer à travers le Web ces éléments de métadonnées.

La description des ressources HTML par les producteurs de ces documents nécessite une approche différente : l'intégration des attributs RDFa dans le *BODY* du document. Cela revient à incorporer dans le document HTML des *tags* MLR en

utilisant le principe des CURI. On peut aussi encourager les utilisateurs à contribuer à la description des ressources par l'ajout des tags à ces dernières. L'utilisation des tags compréhensibles par les machines (*machine tags*) est alors conseillée en vue d'assurer une certaine uniformité en proposant aux utilisateurs des éléments MLR de base qui seront associés aux ressources à décrire ou à commenter. Ces tags sont composés de trois parties : l'espace de nom (*namespace*), la clé et sa valeur :

<b>Espace de nom : Clé = Valeur</b>
-------------------------------------

Exemple :                    **mlr:title= "Introduction au Web sémantique"**

Ainsi, en utilisant le MLR, les balises sémantiques peuvent être intégrées dans des ressources XHTML par les utilisateurs eux-mêmes en autant qu'ils ont les compétences nécessaires. Ces balises seraient ensuite extraites sous forme de triplets RDF en utilisant des transformations GRDDL.

Les sections précédentes nous ont montré que grâce principalement à RDFa et le mécanisme de glanage de ressources (GRDDL), le MLR peut déjà entrer dans le monde du Web sémantique. La visibilité des ressources d'enseignement et d'apprentissage pour leur exploitation et réutilisation est liée à leur indexation. Les normes de métadonnées ont été parmi les premières à avoir atteint un niveau de maturité reconnu dans le domaine de l'éducation (Nilsson, 2008). En dépit des progrès accomplis, plusieurs défis restaient à relever: la nécessité d'associer de la sémantique aux ressources du Web pour en faciliter le traitement par des utilisateurs humains et machines. Les spécifications du MLR lui permettent de relever ces défis et de répondre aux exigences d'interopérabilité avec les schémas

existants tels que Dublin Core ou LOM (Nilsson & Baker, 2008). En utilisant la syntaxe de RDFa, le MLR offre au monde de l'éducation un outil capable de décrire les ressources d'enseignement et d'apprentissage d'une manière compréhensible par les machines. Et en soi, cela est une réponse à ceux qui questionnent l'arrivée d'une autre norme de métadonnées dans un secteur qui n'en manque pas.

## CHAPITRE VI

### CONCLUSION

L'accès à une masse phénoménale de ressources d'information détenues et gérées par différents acteurs publics et privés ne peut plus se concevoir dans un cadre strictement restreint et réservé aux utilisateurs privilégiés et aux producteurs de données. Grâce à l'Internet et particulièrement aux technologies Web, ces ressources font désormais partie du patrimoine informationnel de l'humanité. Les métadonnées sont un élément important pour l'échange, le partage et l'accès à ces ressources. La croissance exponentielle des ressources d'information numériques s'est accompagnée d'une prolifération des schémas de métadonnées, chacun étant conçu pour répondre aux exigences et aux besoins d'une communauté d'utilisateurs déterminée ou d'un secteur d'activité comme celui de l'éducation.

Depuis quelques années, le nombre de schémas de métadonnées est en constante augmentation. Cette diversité des schémas est un des facteurs à la base du problème d'interopérabilité des métadonnées, un problème susceptible d'affecter négativement le repérage, le partage et la réutilisation des ressources.

Face à ce problème d'interopérabilité des métadonnées, il existe en définitive deux solutions (Wüstner et al., 2002). La solution idéale serait l'adoption et l'utilisation par diverses communautés d'un seul et unique schéma de métadonnées. Malheureusement, nous avons vu, tout au long de ce travail, que cette solution est une utopie. Aucun schéma ne peut répondre aux besoins de plusieurs communautés qui, pour des raisons légitimes, n'ont eu d'autre choix que de développer leurs

propres schémas de métadonnées. La diversité des schémas de métadonnées est donc incontournable. Aussi, plusieurs techniques ont été envisagées pour améliorer l'interopérabilité des métadonnées.

La solution la plus fréquente est le *mapping* qui fait appel à différents mécanismes de conversion afin de rendre différents schémas de métadonnées interopérables les uns avec les autres. Malheureusement, les coûts associés à ce mécanisme sont, en général, trop élevés même si l'utilisation des schémas de métadonnées normalisés permet de réduire les coûts de conversion d'un schéma à l'autre. De fait, la normalisation constitue le premier pas vers l'interopérabilité. Le mapping n'est envisageable que dans la mesure où il s'appuie sur des normes, autrement les coûts associés à ce processus seraient exorbitants (figure 6.1). Le coût total comprend les coûts reliés à l'élaboration d'un outil de conversion ainsi que les coûts reliés aux résultats du processus (perte d'information, intervention humaine pour compenser les limites et corriger les erreurs de l'outil de conversion).

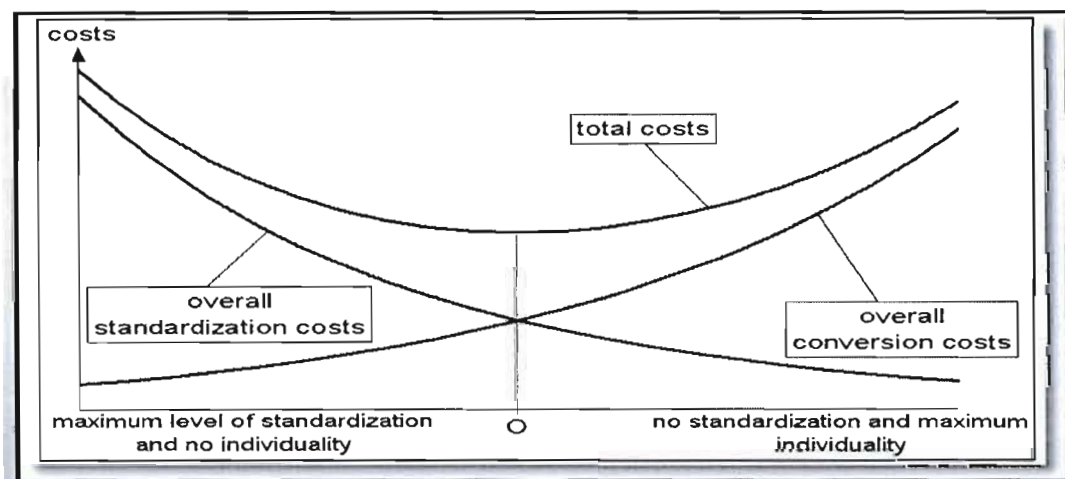


Figure 6.1 : Coûts de normalisation vs coûts de conversion

(Wüstner et al., 2002, p. 55).

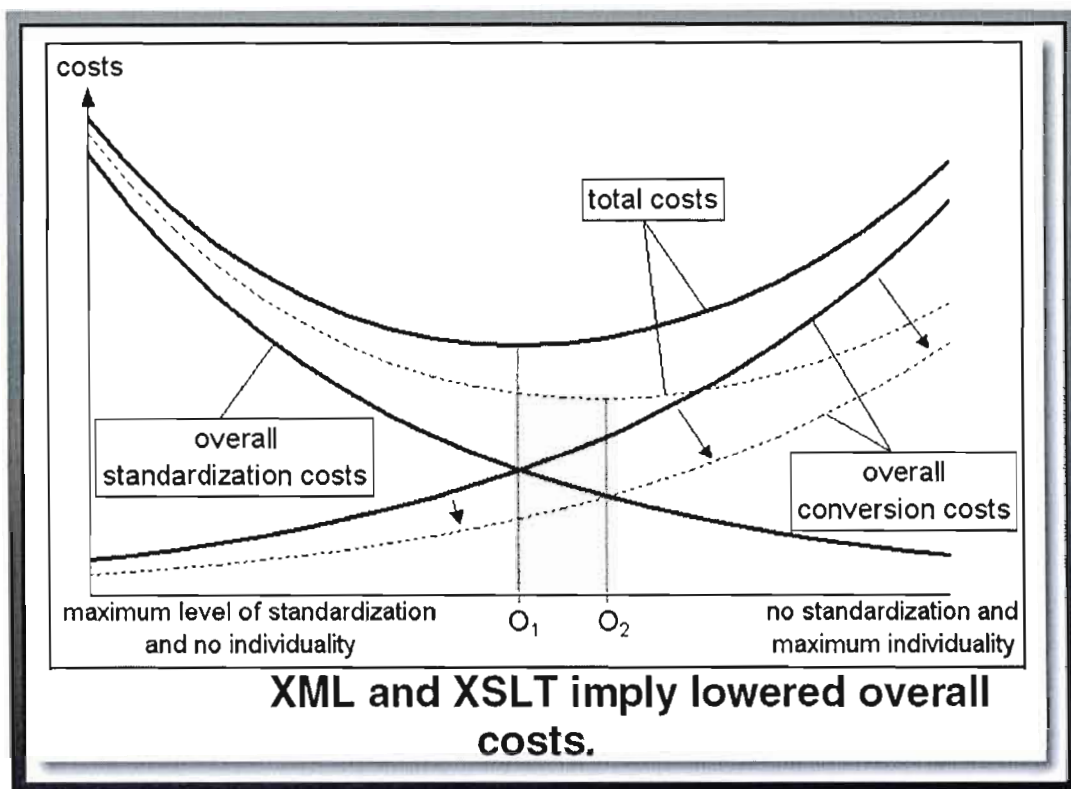


Figure 6.2 : Technologies XML et coûts de conversion (Wüstner et al., 2002, p. 57).

Les technologies XML ont contribué énormément à résoudre le problème d'interopérabilité des métadonnées. XSLT a facilité grandement la conversion et l'échange des métadonnées. En effet, les coûts de production d'une feuille de style XSLT sont moins élevés par rapport aux coûts générés par les méthodes traditionnelles (figure 6.2). Pas étonnant que cette approche soit largement exploitée depuis quelques années dans le monde des bibliothèques pour échanger des données bibliographiques ou pour convertir des métadonnées orientées Web dans les formats MARC et vice-versa.



Au chapitre 1, nous avons distingué les différents niveaux d'interopérabilité. Si l'interopérabilité syntaxique trouve sa solution dans l'utilisation du langage XML, il en est autrement de l'interopérabilité sémantique. XML ne disposant pas de sémantique, il est nécessaire de recourir à d'autres solutions : les technologies du Web sémantique. Grâce à ces technologies, la coexistence de plusieurs schémas de métadonnées n'est pas un obstacle insurmontable à l'interopérabilité sémantique.

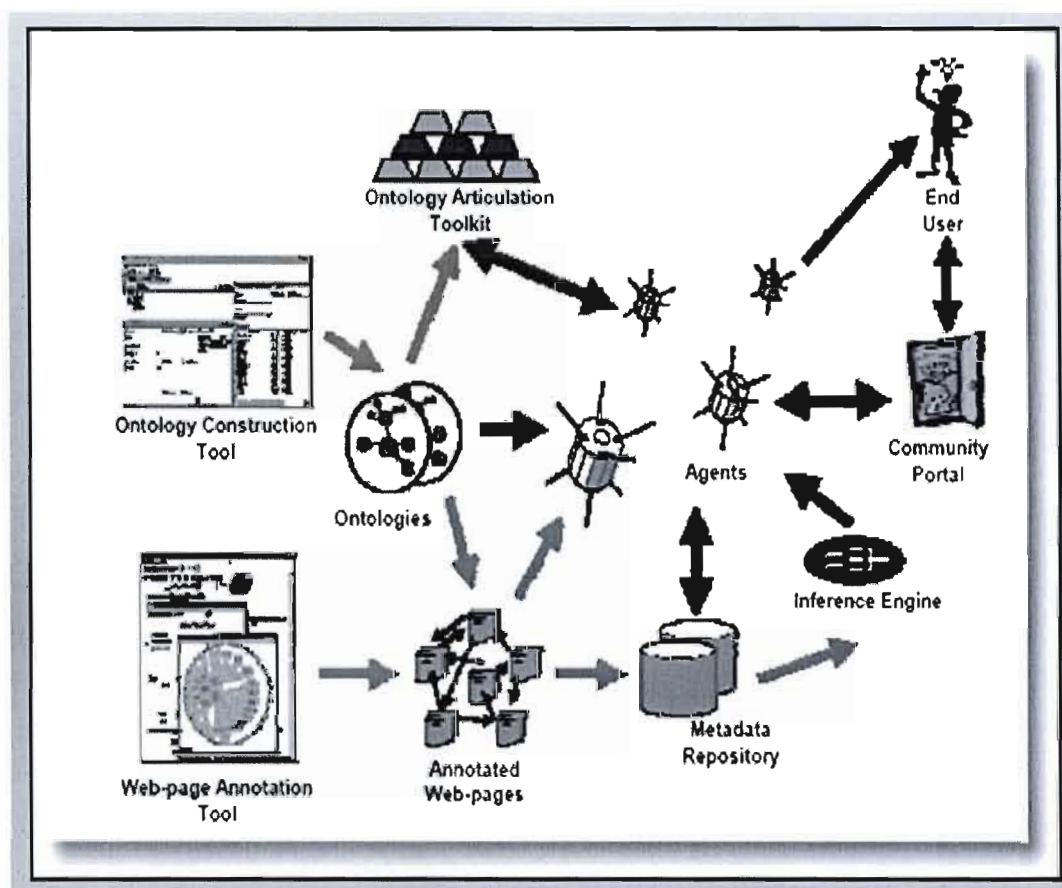


Figure 6.3 : Web sémantique et métadonnées <sup>(75)</sup>

<sup>75</sup> [www.semanticWeb.org](http://www.semanticWeb.org)



L'interopérabilité sémantique dépend de la terminologie, c'est-à-dire de la façon dont les concepts s'expriment dans un vocabulaire contrôlé ayant le statut de référentiel d'un domaine de connaissance ou d'activité et les relations sémantiques entre ces concepts. La figure 6.3 donne une vision du rôle des métadonnées dans le Web sémantique : des ressources Web sont annotées à l'aide des métadonnées issues d'ontologies dont le but est de normaliser la sémantique de ces annotations. Regroupées dans des entrepôts de métadonnées et distribuées sur le Web, ces annotations deviennent exploitables par des agents de recherche d'information. La disponibilité des « référentiels terminologiques » est donc une question cruciale qui n'est pas à sous-estimer. Le chapitre 3 a discuté du rôle important, pour la réalisation du Web sémantique, des ontologies et des langages permettant de les construire et de les mettre en œuvre. En effet, ces vocabulaires permettent aux applications qui communiquent entre elles d'utiliser les mêmes termes avec la même sémantique. Nous avons également souligné que, non seulement l'élaboration, la maintenance et l'évolution de ces instruments terminologiques sont trop coûteuses mais aussi que certains d'entre eux, et en particulier les ontologies, sont hermétiques au commun des mortels. Même si les bénéfices potentiels apportés par ces nouvelles possibilités d'application sont indéniables, il ne semble pas exister beaucoup d'applications réellement utilisées de façon routinière sur le Web. La problématique de la compréhension et de l'adoption de ces technologies du Web sémantique se pose.

Face à la complexité des technologies du Web sémantique, la communauté Web a répondu avec une approche pragmatique qui n'exige pas des changements drastiques dans les façons de faire : ce sont les microformats. Ces microformats sont des conventions permettant d'ajouter des annotations sémantiques aux documents HTML, tout en évitant le recours à de nouveaux langages. Évidemment leur

puissance est très limité par rapport à RDF/OWL et se limitent bien souvent à un domaine précis. Ils constituent néanmoins une étape importante dans la mise en œuvre du Web sémantique. Préconisés par le W3C et inspirés des microformats, les attributs RDFa sont plus puissants et offrent une flexibilité plus grande que les microformats. Incorporés dans une page HTML, les déclarations RDF peuvent être extraites des pages HTML grâce au mécanisme GRDDL et convertis en triplets RDF. Ils peuvent alors constituer des fichiers HTML autonomes et mis à la disposition des applications du Web sémantique tel que les moteurs de recherche. Ce processus de distribution des métadonnées sur le Web présente des avantages indéniables par rapport au modèle centralisé dans lequel les notices de métadonnées sont conservées dans des bases de données à la manière des catalogues de bibliothèques. C'est dans ce contexte que s'inscrit ce mémoire : explorer les voies pour distribuer les notices *MLR* dans le Web. Nous avons proposé un mécanisme de distribution des éléments de métadonnées *MLR* et les rendre disponibles aux applications du Web sémantique. Les attributs RDF (RDFa) sont au centre de cette approche.

Nous avons choisi le *MLR* pour diverses raisons. En particulier, pour mieux répondre aux exigences du Web sémantique, les concepteurs du *MLR* ont privilégié une approche où les éléments *MLR* seraient par essence distribués sous forme de triplets RDF. Notre travail est basée sur le CD3. En mai dernier, une modification du modèle à été proposée. Ce nouveau modèle préserve les caractéristiques du *MLR* présentées ici mais il est mieux adapté à la nécessité d'avoir des éléments distribués sur le Web et de s'adapter aux exigences d'indexation collaborative et distribuée du Web 2.0. Nos travaux ont été faits parallèlement à l'élaboration du *MLR*. À titre d'exemple, l'absence de spécification du *binding XML* pour le *MLR* ne nous a pas

permis de proposer des outils XSLT tel que nous l'aurions souhaité <sup>(76)</sup>. Voilà une perspective de recherche souhaitable à court terme. Tout ce travail n'a évidemment de sens que si des outils tel que *Swoogle* puissent être capables de ramasser ces triplets RDF et de les exploiter. Certes, avec les microformats et les attributs RDFa, nous sommes déjà entrés dans le Web sémantique, mais il est vrai que nous ne pourrions pas nous contenter indéfiniment de ces palliatifs au Web sémantique.

Aujourd'hui, les systèmes de recherche de l'information s'appuient sur un ensemble riche de vocabulaires venant du monde de la bibliothéconomie. Diverses communautés ont défini, à travers les âges, des schémas de métadonnées adaptés à leurs pratiques et dont le Web sémantique ne peut ignorer. L'utilisation des thesaurus et autres langages documentaires non compatibles avec le Web afin d'indexer des documents est une réalité qui ne peut être ignorée. Ces outils constituent un héritage extrêmement riche sur lequel le Web sémantique doit s'appuyer. Vu la complexité du langage de représentation des ontologies, le OWL, il paraît nécessaire de trouver une autre alternative plus facile à utiliser même s'il est moins puissant. Dans l'état actuel de la recherche, et en attendant de trouver une solution efficace pour cacher la complexité du langage OWL, le langage SKOS <sup>(77)</sup> (Miles & Bechhofer, 2008) paraît une alternative crédible pour faire évoluer cet héritage bibliothéconomique dans le Web sémantique. Cependant, la simplicité relative de SKOS n'élimine pas, en définitive, la nécessité de passer un jour aux ontologies et au langage OWL. Le besoin d'outils faciles à utiliser reste un défi de taille pour la recherche sur le Web sémantique, d'autant plus que l'annotation des ressources sera de plus en plus collaborative.

---

<sup>76</sup> Aux dernières nouvelles, le *mapping* RDF est prévu en juillet 2009

<sup>77</sup> Simple knowledge organization system (<http://www.w3.org/2004/02/skos/>)

## BIBLIOGRAPHIE

Adida, B. (2008). hGRDDL: Bridging microformats and RDFa. *Journal of web semantics*, 6, 54-60.

Adida, B., Birbeck, M., McCarron, S., & Pemberton, S. (2008). RDFa in XHTML: Syntax and Processing : A collection of attributes and processing rules for extending XHTML to support RDF. Consulté le 10 février 2008 à l'adresse <http://www.w3.org/TR/rdfa-syntax>

Alesso, P. H., & Smith, C. F. (2005). *Developping Semantic Web Services*: A.K. Peters.

Allsopp, J. (2007). *Microformats: Empowering Your Markup for Web 2.0*. Berkeley: Friends of ED.

Amerouali, Y. (2000). Métadonnées et bibliothèques numériques. Consulté le 12 mai 2006 à l'adresse <http://debora.enssib.fr/LYON1.1.PDF>

Aussenac-Gilles, N. (2007). Introduction au Web sémantique. Consulté le 14 mars 2007 à l'adresse [www.unice.fr/urfist/IMG/ppt/WebSemantique1-2006.ppt](http://www.unice.fr/urfist/IMG/ppt/WebSemantique1-2006.ppt)

Baca, M., Gill, T., Gilliland, A. J., & Woodley, M. S. (2000). Introduction to metadata: pathway to digital information. Consulté le 21 janvier 2004 à l'adresse [http://www.getty.edu/research/conducting\\_research/standards/intrometadata/glossary.html](http://www.getty.edu/research/conducting_research/standards/intrometadata/glossary.html)

Baker, T., & Dekkers, M. (2003a). Identifying Metadata Elements with URIs. *D-Lib Magazine*, 9 (7/8).

Baker, T., & Dekkers, M. (2003b). Identifying Metadata Elements with URIs: the *CORES Resolution*. *D-Lib Magazine*, 9(7/8). Consulté le 20 mai 2007 à l'adresse <http://www.dlib.org/dlib/july03/baker/07baker.html>

Ball, A. (2007). Briefing Paper : the Adobe eXtensible Metadata Platform (XM). Consulté le 25 février 2007 à l'adresse <http://www.ukoln.ac.uk/projects/grand-challenge/papers/xmpBriefing.pdf>

Barbosa, D. (2008). The Taxonomy Folksonomy Cookbook : finding the right recipe for organizing Enterprise Metadata. Consulté le 21 janvier 2008 à l'adresse [http://solutions.dowjones.com/cookbook/ebook\\_sla2008/cookbookebook\\_rwt.pdf](http://solutions.dowjones.com/cookbook/ebook_sla2008/cookbookebook_rwt.pdf)

Berners-Lee, T., Hendler, J., & Lasilla, O. (2001). The Semantic Web : a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, 284(5), 35-43.

Bouquet, P., Serafini, L., & Zanobini, S. (2006). *Bootstrapping semantics on the web : meaning elicitation from schemas*. Papier présenté à la conférence WWW 2006. Consulté le 26 mai 2007 à l'adresse <http://www.dit.unitn.it/~p2p/RelatedWork/Matching/p505-bouquet.pdf>

Bourda, Y. (2008). *Des métadonnées à l'adaptation* (Mémoire présenté en vue de l'obtention de l'habilitation à diriger les recherches). Paris: Université Paris Sud

Bourda, Y., & Delestre, N. (2005a). Améliorer l'interopérabilité des profils d'application du LOM. *Sciences et technologies de l'information et de la communication pour l'éducation et la formation*, 12.

Bourda, Y., & Delestre, N. (2005b). Améliorer l'interopérabilité des profils d'application du LOM. *Revue STICEF*, 12, 15 p.

BSI (2007). *BS 8723-4 :2007 Structured vocabularies for information retrieval – Part 4 : Interoperability between vocabularies*: British Standards Institution (BSI).

Canada, Secrétariat du Conseil du trésor. (2005). *Guide d'élaboration et de mise à jour des vocabulaires contrôlés au gouvernement du Canada*. Consulté le 3 mars 2007 à l'adresse <http://www.tbs-sct.gc.ca/im-gi/mwg-gtm/cvsg-sgvc/docs/2005/vocab/vocab01-fra.asp>.

Cardoso, J. (2007). The Syntactic and the Semantic Web. In *Semantic Web Services : Theory, Tools and Applications* (pp. 1-23). Lond: IGI publishing.

Caussanel, J., Cahier, J.-P., Zacklad, M., & Charlet, J. (2002). Les Topic Maps sont-ils un bon candidat pour l'ingénierie du Web sémantique ? Université de technologie de Troyes, Laboratoire "Technologies de la Coopération pour l'innovation et le changement organisationnel". Consulté le 15 juin 2004 à l'adresse <http://cahier.tech-cico.fr/publi/caussanel-IC2002.pdf>

Cearley, D. W., Andrews, W., & Gall, N. (2007). *Finding and exploiting value in semantic technologies on the Web*. 10p. (Gartner Research, ID G00148725)

Chawk, M. (2004). *L'usage des métadonnées dans la description et la recherche des ressources pédagogiques sur le WEB* Papier présenté à la conférence Colloque TICE Méditerranée. Consulté le, Disponible à l'adresse [http://atrium.unice.fr/nte/colloque/communication\\_fichiers/16-chawk-ben-abadallah.pdf](http://atrium.unice.fr/nte/colloque/communication_fichiers/16-chawk-ben-abadallah.pdf)

Cherhal, E. (2004). *Présentation des standards : le Dublin Core*. Papier présenté à la conférence Journée "Indexation des ressources pédagogiques". Consulté le 2 mars 2007. Disponible à l'adresse <http://www.mathdoc.fr/Publications/Textes/cherhal-enssibdc.pdf>

Cousins, S., & Sanders, A. (2006). Incorporating a virtual union catalogue into the wider information environment through the application of middleware : Interoperability issues in cross-database access. *Journal of documentation*, 62 (1), 120-144.

Daconta, M. C., Obrst, L. J., & Smith, K. T. (2003). *The Semantic Web : A Guide to the Future of XML, Web Services and Knowledge Management*. Indianapolis: Wiley.

Davis, M. (2007). *Semantic wave 2007 : industry roadmap to web 3.0. tutorial* Papier présenté à la Semantic technology conference 2007.

Degoulet, P., Fieschi, M., & Attali, C. (1997). Les enjeux de l'interopérabilité sémantique dans les systèmes d'information de santé. *Informatique et Santé*, 9, 203-212.

Delestre, N., & Bourda, Y. (2005). Utilisation de la norme ISO11179 pour améliorer l'interopérabilité entre les différents schémas de métadonnées pédagogiques. Disponible à l'adresse [http://hal.archives-ouvertes.fr/docs/00/02/75/65/PDF/Delestre\\_Bourda.pdf](http://hal.archives-ouvertes.fr/docs/00/02/75/65/PDF/Delestre_Bourda.pdf)

Dream'Orange (2007). Les microformats : à la croisée du web 2.0 et du web sémantique. Consulté le 2008-03-20 à l'adresse <http://www.dreamorange.fr/resources/dossiers/microformats/microformats.pdf>

Duval, E., & Hodgins, W. (2002). Metadata principles and practicalities. *D-Lib Magazine*, 8 (4). Disponible à l'adresse <http://www.dlib.org/dlib/april02/weibel/04weibel.html>

EICTA (2004). EICTA interoperability white paper. Disponible à l'adresse [http://www.eic-community.org/index.php?option=com\\_docman&task=doc\\_view&gid=15&Itemid=351](http://www.eic-community.org/index.php?option=com_docman&task=doc_view&gid=15&Itemid=351)

Euzenat, J. (2003). Qu'est-ce que le web sémantique ? Consulté le 13 avril 2004 à l'adresse <ftp://ftp.inrialpes.fr/pub/exmo/publications/euzenat2003d.pdf>

Feigenbaum, L., & Torres, E. (2006). A meaningful web for humans and machines : part 1 : how humans can share the wealth of the web. Consulté le 1er novembre 2006 à l'adresse <http://www.ibm.com/developerworks/xml/library/wa-hmach1/index.html>

Foster, R. N. (1986). *Innovation: The Attacker's Advantage*. New York: Summit Books.

Friesen, N. (2004). The International Learning Object Metadata Survey. *The International Review of Research in Open and Distance Learning* Consulté le 15 mars 2007 à l'adresse <http://www.irrodl.org/content/v5.3/technote5.html>

Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1987). The vocabulary problem in human-system communication. *Communications of the ACM*, 30 (11), 964-968.

Gandon, F. (2008). Profile for latest GRDDL transformation for RDFa. Consulté le 10 septembre 2008 à l'adresse <http://ns.inria.fr/grddl/rdfa/>

Geroimenko, V. (2006). The Concept and Architecture of the Semantic Web. *In* V. Geroimenko & C. Chaomei (Eds.), *Visualizing the Semantic web : XML-based Internet and Information Visualization* (pp. 3-17). London: Springer-Verlag.

Ghafour, S. A. (2003). *Méthodes et outils pour l'intégration des ontologies*. Lyon: INSA. Disponible à l'adresse <http://liris.cnrs.fr/Documents/Liris-1747.pdf>

Golbreich, C., Dameron, O., Gibaud, B., & Burgun, A. (2003). Standards et ontologies biomédicales pour un web sémantique. Consulté le 3 août 2007 à l'adresse [https://lists.lirmm.fr/wws/d\\_read/ontobio/JourneesJuin2003/C-Golbreich.pdf](https://lists.lirmm.fr/wws/d_read/ontobio/JourneesJuin2003/C-Golbreich.pdf).

Graf, A. (2007). RDFa vs. Microformats. *DERI technical report 2007-04-10* Consulté le 24 avril 2008 à l'adresse [http://www.sti-innsbruck.at/fileadmin/documents/technical\\_report/html\\_metadata/RDFaVsMicroformats.pdf](http://www.sti-innsbruck.at/fileadmin/documents/technical_report/html_metadata/RDFaVsMicroformats.pdf)

Gruber, R. T. (1993). A translation approach portable ontology specifications. *Knowledge acquisition*, 5 (2), 199-220.

Guay, P. J. (2007). Pour mieux décrire les ressources d'enseignement et d'apprentissage. *Clic: Bulletin collégial des technologies de l'information et des communications*, 64, avril 2007. Disponible à l'adresse <http://clic.ntic.org/cgi-bin/aff.pl?page=article&id=2036>

Halpin, H., & Davis, I. (2007). Introduction à GRDDL. Consulté le 9 janvier 2008 à l'adresse [www.w3.org/TR/2007/NOTE-grddl-primer-20070628](http://www.w3.org/TR/2007/NOTE-grddl-primer-20070628)

Hazaël-Massieux, D. (2008). GRDDL service : extracting RDF from XHTML/XML using GRDDL. Consulté le 30 juillet 2008 à l'adresse <http://www.w3.org/2007/08/grddl/>

Heery, R., & Patel, M. (2000). Application profiles : mixing and matching metadata schemas. *Ariadne*, no 25. Disponible à l'adresse <http://www.ariadne.ac.uk/issue25/app-profiles/>

Heiler, S. (1995). Semantic interoperability. *ACM Computing Survey*, 27 (2), 271-273.

Hori, M. (2003). Semantic Annotation for Web Content Adaptation. *In* D. Fensel, J. Hendler, H. Lieberman & W. Wahlster (Eds.), *Spinning the Semantic Web : Bringing the World Wide Web to Its Full Potential*: MIT Press.



Hudon, Y. (2001). *L'adaptabilité culturelle et linguistique : un enjeu capital pour l'interopérabilité pour les infrastructures et une opportunité d'affaires sur Internet*. Consulté le 15 mars 2003. Disponible à l'adresse [http://www.msg.gouv.qc.ca/fr/publications/enligne/societe/francais/adaptation\\_culturelle.pdf](http://www.msg.gouv.qc.ca/fr/publications/enligne/societe/francais/adaptation_culturelle.pdf)

INTEROP (2003). INTEROP : Interoperability research for networked enterprises applications and software. Disponible sur le site <http://www.interop-noe.org>

iProspect (2004). iProspect's search engine user attitudes survey results. Consulté le 3 juillet 2006 à l'adresse <http://www.iprospect.com/premiumPDFs/iProspectSurveyComplete.pdf>

ISO/CEI JTC1 SC36 (2004). Final Report on the "International LOM Survey". Consulté le 10 juin 2007 à l'adresse <http://staff.uow.edu.au/eteaching/str/metadata/intlLOmsurvey.pdf>. Document SC36 N-0871

ISO/CEI JTC1 SC36 (2008a). Information technology - Learning, education and training - Metadata for learning resources - Part 1 : Framework, ISO/IEC CD3 19788-1.

ISO/CEI JTC1 SC36 (2008b). Information technology - Learning, education and training - Metadata for learning resources - Part 2 : Core elements, ISO/IEC CD3 19788-2.

ISO/CEI JTC1 SC36 (2008c). Working document resulting from the CD3 MLR-1 BRM Paris. Consulté le 10 décembre 2008 à l'adresse [http://isotc.iso.org/livelink/livelink/fetch/2000/2122/327993/806742/1056984/36N1732\\_Resolutions\\_of\\_the\\_MLR-1\\_CD3\\_BRM\\_in\\_Paris\\_France.pdf?nodeid=7729818&vernum=0](http://isotc.iso.org/livelink/livelink/fetch/2000/2122/327993/806742/1056984/36N1732_Resolutions_of_the_MLR-1_CD3_BRM_in_Paris_France.pdf?nodeid=7729818&vernum=0)

ISO/CEI JTC1 SC36 (2009). Sneak Peek FCD for ISO/IEC 19788-1, Metadata for Learning Resources, Part 1 : Framework, (2009).

Johannis, P. (2005). *Documentation sur les éléments de données dans les organismes statistiques*. Ottawa: Statistique Canada.

Khare, R., & Celik, T. (2006). Microformats : a pragmatic path to the semantic web. *CommerceNet Labs Technical Report, 06-01*. Consulté le 11 mai 2007 à l'adresse <http://wiki.commerce.net/images/e/ea/CN-TR-06-01.pdf>

La Passardière, B. d., & Jarraud, P. (2004). ManUel, un profil d'application du LOM pour C@mpusSciences. *Sciences et technologies de l'information et de la communication pour l'éducation et la formation, 11*.

Lacy, L. W. (2005). *OWL : Representing information using the Web ontology language*. Victoria, B.C: Trafford.



Landry, P. (2006). Multilinguisme et langages documentaires : le projet MACS en contexte européen. *Documentation et bibliothèques*, 52 (2), 121-129

Landry, P., Zumer, M., & Clavel-Merrin, G. (2006). *Report on Cross-Language Subject Access Options*. s.l.: The European library. Disponible à l'adresse <http://www.nuk.uni-lj.si/telmemor/docs/D3.4-Cross-language-access.pdf>

Larher, T., & Milbeau, K. (1999). Les métadonnées. *Internet au quotidien*, octobre 1977, 37-39.

Laublet, P., Reynaud, C., & Charlet, J. (2003). *Sur quelques aspects du web sémantique*. Actes des des deuxièmes Assises nationales du GdR. Disponible à l'adresse <http://enssibal.enssib.fr/autres-sites/RTP/websemanique/articles/03-WebSemantique.pdf>.

Menon, B. (2004). Web sémantique et traitement automatique des langues. Consulté le 10 mars 2006 à l'adresse <http://www.i-expo.net/documents/actes2004/i3/BrunoMenon.pdf>

Miles, A., & Bechhofer, S. (2008). SKOS Simple Knowledge Organization System Reference. Consulté le 2008-08-14 à l'adresse <http://www.w3.org/TR/2008/WD-skos-reference-20080609/>

Mizoguchi, R., & Bourdeau, J. (2004). Le rôle de l'ingénierie ontologique dans le domaine des EIAH. *Revue STICEF*, 11.

Moen, W. E. (2001). *Mapping the Interoperability Landscape for Networked Information Retrieval*. Papier présenté à la conférence JCDL 2001. Disponible à l'adresse <http://www.unt.edu/wmoen/publications/MapInteropJCDLFinal.pdf>

Moen, W. E. (2004). *Metadata Interaction, Integration, and interoperability*. Papier présenté à la conférence NISO Workshop « Metadata Practices on the Cutting Edge », 2004. Disponible à l'adresse [www.unt.edu/wmoen/presentations/MD\\_Interoperability\\_wem\\_June2005.ppt](http://www.unt.edu/wmoen/presentations/MD_Interoperability_wem_June2005.ppt)

Morel-Pair, C. (2005). *Panorama - des métadonnées pour les ressources électroniques*. Papier présenté à la conférence Ateliers des réseaux de la documentation française. Disponible à l'adresse [http://hal.archives-ouvertes.fr/docs/00/04/04/73/PDF/Metas\\_panorama\\_CMO.pdf](http://hal.archives-ouvertes.fr/docs/00/04/04/73/PDF/Metas_panorama_CMO.pdf)

Nilsson, M. (2008). Harmonization of Metadata and Standards. *Learning objects, Metadata and Standards*. Disponible à l'adresse <http://ariadne.cs.kuleuven.be/lomi/images/5/52/D4.7-prolearn.pdf>

Nilsson, M., & Baker, T. (2008). Requirements for ISO MLR interoperability. Consulté le 20 juillet 2008 à l'adresse [http://isotc.iso.org/livelink/livelink/fetch/2000/2122/327993/806742/4920436/4920313/WG4\\_N0255\\_Requirements\\_for\\_ISO\\_MLR\\_Interoperability\\_--Jeju\\_Presentation\\_of\\_Document\\_WG4\\_N0238.pdf?nodeid=7223540&vernum=0](http://isotc.iso.org/livelink/livelink/fetch/2000/2122/327993/806742/4920436/4920313/WG4_N0255_Requirements_for_ISO_MLR_Interoperability_--Jeju_Presentation_of_Document_WG4_N0238.pdf?nodeid=7223540&vernum=0)

Nilsson, M., Powell, A., & Baker, T. (2007). DCMI Abstract Model. Consulté le 2008-05-20 à l'adresse <http://dublincore.org/documents/abstract-model/>

NISO (2004). Understanding metadata, Disponible à l'adresse [www.niso.org/publications/press/UnderstandingMetadata.pdf](http://www.niso.org/publications/press/UnderstandingMetadata.pdf)

Nogueras-Iso, J., Zarazaga-Soria, F. J., Lacasta, J., Béjar, R., & Muro-Medrano, P. R. (2004). Metadata standard interoperability: application in the geographic information domain. *Computers, Environment and Urban Systems*, 28, 611-634.

Obrst, L. (2003). *Ontologies for semantically interoperable systems*. In : Proceedings of the twelfth international conference on Information and knowledge management, 2003, p. 366-369

OCLC (1995). *The OCLC Online Computer Library Center Office of Research and The National Center for Supercomputing Applications announces an intensive workshop focused on the description of network resources : The Essential Elements of Network Object Description*. Consulté le 1er mai 2008 à l'adresse <http://dublincore.org/workshops/dc1/general.shtml>

Oubahssi, L. (2005). *Conception de plates-formes logicielles pour la formation à distance, présentant des propriétés d'adaptabilité à différentes catégories d'utilisateurs et d'interopérabilité avec d'autres environnements logiciels.*, Université René Descartes - Paris V, Paris.

Paskin, N. (2004). The <indec> Data Dictionary. Consulté le 5 septembre 2005 à l'adresse [www.doi.org/topics/041202BICNISO.ppt](http://www.doi.org/topics/041202BICNISO.ppt)

Porquet, C. (2005). Une introduction au web sémantique. Consulté le 21 janvier 2006 à l'adresse [www.greyc.ensicaen.fr/~chris/Cours\\_ws\\_cp\\_2005.htm](http://www.greyc.ensicaen.fr/~chris/Cours_ws_cp_2005.htm)

RCIP (2002). Normes relatives aux métadonnées pour le catalogage des collections de musée. Consulté le 20 juillet 2006 à l'adresse [www.rcip.gc.ca/Francais/normes/metadonnees\\_correspondances.html](http://www.rcip.gc.ca/Francais/normes/metadonnees_correspondances.html)

Rebai, I., & Labat, J.-M. (2006). *Un outil d'aide à la création de profils d'application*. Papier présenté à la conférence TICE 2006. Consulté le 2 mars 2007. Disponible à l'adresse [http://www.math-info.univ-paris5.fr/~rebi/data/recherche/TICE2006\\_Rebai\\_Labat.pdf](http://www.math-info.univ-paris5.fr/~rebi/data/recherche/TICE2006_Rebai_Labat.pdf)

Roussel, C., Goettmann, E., & Nicolas, Y. (2007). Métadonnées : faut-il parier sur RDF (Resource Description Framework)? Consulté le 12 novembre 2007 à l'adresse [http://artist.inist.fr/article.php3?id\\_article=406](http://artist.inist.fr/article.php3?id_article=406)

Schapranow, M.-P. (2006). Microformats: A XML document immanent way for semantic annotations of web contents. Consulté le 15 novembre 2008 à l'adresse <http://www.myhpi.de/~schapran/pke/Microformats - A XML-document-immanent way for semantic annotations of web contents.pdf>

SCTIC (2002). *Les normes et standards de la formation en ligne : état des lieux et enjeux*. Montréal: PROFETIC. Disponible à l'adresse <http://www.profetic.org/file/norm-0210-d-RAPPORT.pdf>

SDTIC (2007). *Indexation et visibilité des ressources pédagogiques : un enjeu pour la valorisation du patrimoine pédagogique des établissements d'enseignement supérieur*. Papier présenté à la conférence Séminaire Indexation et visibilité des ressources, 26 et 27 mars 2007 Paris. Disponible à l'adresse [http://www.educnet.education.fr/chrgt/seminaire\\_indexation\\_2007.pdf](http://www.educnet.education.fr/chrgt/seminaire_indexation_2007.pdf)

Sidhom, S. (2005). Catalogues électroniques multimédias : performances. Consulté le 24 avril 2008 à l'adresse [www.loria.fr/~ssidhom/COURS\\_0506/ue303\\_M2-istie/UE303c\\_5\\_M2istie.ppt](http://www.loria.fr/~ssidhom/COURS_0506/ue303_M2-istie/UE303c_5_M2istie.ppt)

Smith, G. (2008). *Tagging : people-powered metadata for the social web*. Berkeley: New Riders.

Spivack, N. (2006). Minding the planet : the meaning and future of the semantic web. Consulté le 20 décembre 2006 à l'adresse [http://novaspivack.typepad.com/nova\\_spivacks\\_weblog/2006/11/minding\\_the\\_pla.html](http://novaspivack.typepad.com/nova_spivacks_weblog/2006/11/minding_the_pla.html)

Spivack, N. (2007a). Beyond keyword and natural language search. Consulté le 15 janvier 2008 à l'adresse [http://novaspivack.typepad.com/nova\\_spivacks\\_weblog/2007/03/beyond\\_keyword\\_.html](http://novaspivack.typepad.com/nova_spivacks_weblog/2007/03/beyond_keyword_.html)

Spivack, N. (2007b). How the WebOS evolves ?". Consulté le 9 février 2007 à l'adresse <http://novaspivack.typepad.com/RadarNetworksTowardsAWebOS.jpg>

Spivack, N. (2007c). A visual timeline of the past, present and future of the web. Consulté le 10 janvier 2008 à l'adresse [http://novaspivack.typepad.com/nova\\_spivacks\\_weblog/2007/02/steps\\_towards\\_a.html](http://novaspivack.typepad.com/nova_spivacks_weblog/2007/02/steps_towards_a.html)

St. Pierre, M., & LaPlant, W. P. (1998). Issues in crosswalking content metadata standards, Disponible sur le site [www.niso.org/press/whitepapers/crswalk.html](http://www.niso.org/press/whitepapers/crswalk.html)

Ta Tuan, A. (2005). *Web sémantique et réseaux sociaux - construction d'une mémoire collective par recommandations mutuelles et (re-)présentations.*, École nationale supérieure des télécommunications, Paris.

Taniar, D., & Rahayu, J. W. (2006). Metadata and ontology based semantic web mining. *In Web Semantics and Ontology*, IGI publishing. p. 259-296.

Tolk, A. (2006). *What comes after the semantic Web - PADS implications for the dynamic web*. Proceedings of the 20th Workshop on Principles of Advanced and Distributed Simulation, 2006. p. 55

Vézina, M.-H. (2007). Environnements numériques. Université de Montréal. Document non publié.

Weibel, S., Miller, E., Godby, J., & Daniel, R. (1995). OCLC/NCSA Metadata Workshop Report. Consulté le 12 janvier 2007 à l'adresse <http://dublincore.org/workshops/dc1/report.shtml>

Woodley, M. S. (2005). DCMI Glossary. Consulté le 2008-01-12 à l'adresse <http://www.dublincore.org/documents/usageguide/glossary.shtml>

Wüstner, E., Hotzel, T., & Buxmann, P. (2002). *Converting business documents : a classification of problems and solutions using XML/XSLT*. Papier présenté à la 4th IEEE international workshop on advanced issues of E-commerce and Web-base information systems, 2002, p. 54-61.

Yonnet, P. (2007). L'autre web sémantique : le web sémantique et les systèmes de recherche d'information. Disponible à l'adresse <http://www.webmaster-hub.com/publication/L-autre-semantique-Le-Web.html>

Zen, M. L., & Chan, L. M. (2006a). Metadata Interoperability and Standardization : a study of methodology, part I : achieving interoperability at the schema level. *D-Lib Magazine*, 12 (6). Disponible à l'adresse <http://www.dlib.org/dlib/june06/chan/06chan.html>

Zen, M. L., & Chan, L. M. (2006b). Metadata Interoperability and Standardization : a study of methodology, part II : achieving interoperability at the record and repository levels. *D-Lib Magazine*, 12 (6). Disponible à l'adresse <http://www.dlib.org/dlib/june06/chan/06chan.html>

## ANNEXE 1

### ÉLÉMENTS DE BASE DU METADATA FOR LEARNING RESOURCES

URI: ISO_IEC_19788-2:2008					
ISO/IEC 19788-2 Clause 5 Term ID	ISO French (fra')				
	Term	Source	G	Definition	Source
(1)	(2)	(3)	(4)	(5)	(6)
DES0010	Identifiant de la ressource	ISO_IEC_19788-2:2008	01	référence stable et non ambiguë à la ressource	ISO_IEC_19788-2:2008
DES0020	Titre	ISO_IEC_19788-2:2008	01	nom attribué à la ressource	ISO_IEC_19788-2:2008
DES0030	Langue de la ressource	ISO_IEC_19788-2:2008	02	langue(s) utilisée(s) dans la ressource pour communiquer avec l'utilisateur prévu	ISO_IEC_19788-2:2008
DES0040	Description de la ressource	ISO_IEC_19788-2:2008	02	énoncé du contenu de la ressource	ISO_IEC_19788-2:2008
DES0050	Identifiant de l'enregistrement MLR	ISO_IEC_19788-2:2008	01	référence stable et non ambiguë à l'enregistrement MLR	ISO_IEC_19788-2:2008
DES0060	Répertoire d'origine de l'enregistrement MLR	ISO_IEC_19788-2:2008	01	identification du répertoire ayant créé l'enregistrement MLR d'origine	ISO_IEC_19788-2:2008
DES0070	Dernière mise à jour de l'enregistrement MLR	ISO_IEC_19788-2:2008	02	date de la plus récente mise à jour de l'enregistrement MLR	ISO_IEC_19788-2:2008
DES0080	Approche pédagogique	ISO_IEC_19788-2:2008	02	utilisation recommandée de la ressource pour la transmission des connaissances, attitudes ou habiletés	ISO_IEC_19788-2:2008
DES0090	Contribution à la ressource	ISO_IEC_19788-2:2008	02	information sur la personne ou l'organisation responsable du développement de la ressource	ISO_IEC_19788-2:2008
DES0091	Rôle joué dans la création de la ressource	ISO_IEC_19788-2:2008	01	type de contribution ou de responsabilité	ISO_IEC_19788-2:2008

DES0092	Personne participant à la création de la ressource	ISO_IEC_19788-2:2008	02	information concernant la personne ou l'organisation ayant contribué à la ressource	ISO_IEC_19788-2:2008
DES0093	Date de participation à la ressourceHumain	ISO_IEC_19788-2:2008	02	date associée à la contribution à la ressource	ISO_IEC_19788-2:2008
DES0100	Instance	ISO_IEC_19788-2:2008	02	expression physique ou électronique de la ressource	ISO_IEC_19788-2:2008
DES0101	Format	ISO_IEC_19788-2:2008	01	expression de la nature électronique, numérique ou physique de la ressource	ISO_IEC_19788-2:2008
DES0102	Taille	ISO_IEC_19788-2:2008	02	taille de la ressource numérique en octets	ISO_IEC_19788-2:2008
DES0103	Localisation	ISO_IEC_19788-2:2008	02	chaîne utilisée pour accéder à la ressource	ISO_IEC_19788-2:2008
DES0110	Droits	ISO_IEC_19788-2:2008	01	description des droits	ISO_IEC_19788-2:2008
DES0111	Enoncé des droits	ISO_IEC_19788-2:2008	01	énoncé textuel des droits	ISO_IEC_19788-2:2008
DES0112	Source des droits	ISO_IEC_19788-2:2008	02	source de l'énoncé des droits	ISO_IEC_19788-2:2008